EUKARYOTIC PLANKTON DIVERSITY ACROSS CONTRASTING SUBTROPICAL AND SUBANTARCTIC OCEANIC WATERS, EAST OF NEW ZEALAND, SW PACIFIC

Gutiérrez-Rodríguez, A., Lopes-Dos Santos, A., Vaulot, D., Not, F., Probert, I., Fernandez D., Bilewitch, J., Hulston, D., Maas, E., Pinkerton, M., Nodder, S.D.

To be submitted to New Zealand Journal of Marine and Freshwater Research, Scientific Reports, Environmental Microbiology Reports (?), Deep Sea Research I

1-INTRODUCTION

**[1] General oceanic microbial diversity/ecology blablabla context** - *Expendable*

Microbial communities represent the majority of marine biomass and constitute a pivotal component of marine ecosystems. Being at the base of the food-web, the marine protist including phototrophic, heterotrophic and mixotrophic single-celled eukaryotes sustain virtually all life and have key roles in the functioning of marine ecosystems (). Phytoplankton is responsible of 50% of global primary productivity (Field et al. 1998). Most of this primary production is consumed and processed by heterotrophic protist (i.e. microzooplankton) before becoming available for larger zooplankton and higher trophic levels (Calbet and Landry, 2004; Saiz and Calbet, 2005; Zeldis and Decima 2019). From a biogeochemical perspective, the microbial production, consumption and remineralization of organic matter is at the core of global biogeochemical cycles including the nitrogen and carbon cycle, and regulates ocean’s capacity to sequester atmospheric CO2 via the biological pump (Turner et al., 2010; Boyd et al., 2019).

The trophic and biogeochemical processes driven by microbial communities are influenced by their taxonomic composition, which is tightly coupled to physico-chemical conditions.. With increasing evidence of climate change on ocean physical conditions (e.g. warming, stratification and reduced mixing)(ref) and flow-on effects on marine ecosystems (ref) it becomes imperative to better characterize the biogeography of microbial communities and their relationship to oceanographic conditions.

However, the high cell abundance, diversity and dynamic nature of microbial communities has precluded a robust characterization of species composition and distributional patterns across contrasting water masses and oceanographic gradients at relevant temporal and spatial scales. The advent of high-throughput Next Generation Sequencing (NGS) technology and development of novel molecular techniques have lead to cost-effective approaches like DNA metabarcoding that allows assessing plankton community biodiversity and functioning over a wide range of temporal to spatial scales (e.g., local (estuaries) to global (e.g., Tara, Guidi et al.) examples) with unprecedented taxonomic coverage and resolution. Biogeographical distribution across large oceanic regions and global oceanographic boundaries/features (e.g. subtropical fronts) are yet to be characterized using these novel approaches. The southwest Pacific is arguably one of the oceanic regions that remains less studied likely due to its extension and remoteness.

SW Pacific surface waters east of New Zealand comprise Subtropical (STW) and Subantarctic (SAW) water masses, separated by the Subtropical Front (STF). These water masses have contrasting physico-chemical and biological characteristics (Heath, 1985; Bradford-Grieve et al., 1999; Boyd et al., 1999; Chiswell et al., 2015). To the east of New Zealand, STW are saltier and warmer surface waters than SAW and are where phytoplankton production is considered to be limited by nitrogen (Zentara and Kamykowski 1981). North of New Zealand, STW is fully oligotrophic (low macro- and micronutrients) with pervasive nitrogen-fixation by diazotrophs (Law et al., 2012; Ellwood et al., 2018). The STF is a dynamic zone, characterized by strong temperature and salinity gradients with high levels of vertical and lateral mixing of high iron STW and macro-nutrient-rich SAW (Chiswell 2001), which consequently leads to regionally elevated annual net primary production (Murphy et al., 2001; Pinkerton et al., 2005). SAW south of the STF are fresher and colder waters where iron is the primary limiting nutrient for phytoplankton growth (Boyd et al., 1999). East of New Zealand SAW are considered high-nutrient, low-chlorophyll, low-silicate waters (HNLC-LSi) where in addition to iron, silicate and light can become limiting at times (Dugdale et al; 1995; Banse, 1996; Boyd et al., 1999). These conditions are typically associated with SAW north of the Subantarctic Front (SAF), which is an area commonly referred to as the Subantarctic Zone (Trull et al. 2001) or the Subantarctic Water Ring (Longhurst, 2007). Further south, increasing dissolved silica availability south of the SAF shifts the Polar Frontal Zone to ‘standard’ HNLC conditions between the SAF and Polar Front (Rigual-Hernandez et al 2015).

Several studies have characterized microbial communities composition in STW and SAW east of New Zealand, using microscopy (Chang et al., 1998), pigment (Delizzo et al., 2009) and flow-cytometry (Hall et al., 2001). These regional studies have focused mainly on the STF zone (Chang et al., 1998; Hall et al., 1999) and coastal communities (Chang et al., 2003; Hall et al., 2006), while studies analyzing wider phytoplankton distribution across ST and SA waters have targeted specific groups such as coccolitophores (e.g. Saavedra-Pellitero et al., 2014; Chang et al., 2016). More process-oriented studies have also provided partial information on phytoplankton composition in SA and ST waters east of New Zealand (Peloquin et al., 2011; Ellwood et al., N-Cycle; Spring Bloom – Twinning et al., 2014; Chiswell et al., 2019; Gutierrez-Rodriguez et al., 2020). A recent study using DNA metabarcoding assessed microbial diversity from tropical to polar latitudes in the SW Pacific region focusing on alpha diversity patterns in relation to physico-chemical gradients and oceanographic features (Raes et al., 2018). However, the taxonomic composition of protistan communities associated with the contrasting water masses in the SW Pacific and across major oceanographic fronts that separate them is still lacking.

The aims of this study are 1) to characterize the diversity of protistan communities in ST and SA waters east of New Zealand and across the major frontal zones - the STF and the Subantarctic Front (SAF) - that delimit these water masses and 2) to investigate the distributional patterns of main protistan taxonomic groups and species in the context of physical and chemical variability across and within water masses. Specifically, we want to know how (dis-)similar are protist communities in ST and SA waters? At which taxonomic level do these differences emerge? What are the main environmental factors responsible for these differences? Which are the main taxonomic groups associated with each water mass and their environmental preferences? To do so we have applied DNA metabarcoding analysis (18S rRNA) to >450 samples collected during 12 oceanographic voyages conducted over several years (2009-2017) and seasons across ST and SA waters east of New Zealand. Our results showed the overall dominance of dinoflagellates and chlorophytes, with consistent differences in the taxonomic composition between water masses emerging at the class and finer taxonomic levels. Analysis of intraspecific diversity revealed differences in the distribution of ASVs of the same species suggesting the presence of different ecotypes in some cases (e.g. Chloroparvula pacifica; Phaeocystis antarctica) and current taxonomic gaps within certain groups that remained to be characterized (e.g. Pelagophyceae\_XXX; Dinophyceae\_XXX).

Oceanographic boundaries - Microbial oceanic provinces – Biomes - Biogeography –

**[4] Aims and questions addressed by this study**

In this study, we aim to:

1. characterize the diversity of protistan plankton diversity across ST and SA waters and in the major frontal zones (STF and SAF) that delimits these water masses,
2. investigate the variability of community composition and richness in relation to environmental factors; and
3. determine the distribution patterns of main protistan taxonomic groups (division/class level) and species (ASV) across environmental gradients across and within water masses.

Questions:

* How (dis-)similar are protist communities in ST and SA waters? At which taxonomic level do these differences emerge (e.g. is the composition similar at division level, class level, or genus level)? Which are the main taxonomic groups characteristic of each water mass? Can we identify species indicators of different water masses?
* What are the main environmental factors responsible for the changes in species richness and community composition as determined using DNA meta-barcoding?
* What are the environmental preferences of the main protistan groups? How do they respond to environmental variability? Do classes or species present across ST and SA waters respond similarly to environmental drivers (e.g. temperature, nitrate, mixing) in both water masses?
* How does community composition respond to mixing associated within the STF zone? Is diversity stimulated by enhanced physico-chemical heterogeneity expected in the STF zone? Can we identify classes/species with preference for the STF zone?

2-METHODS

* 1. Study area and sample collection

Seawater samples and data were collected during 12 cruises conducted in SW Pacific waters east of New Zealand between 2009 and 2017 (Fig. 1). The dataset covered a 100 stations distributed between 54.3 and 33.4 S with seawater samples (n = 479) being collected from sea surface down to 2000 m and during spring, summer and autumn periods (Table 1). The number of DNA samples from STW were 2-fold higher than those from SAW and STF (Table 2). This difference was mainly due to the large number of samples from Spring Bloom II voyage (TAN1212) (Fig. Suppl. 1). s[Seasonal coverage was biased against winter with most samples collected during spring, summer and autumn periods. The seasonanlity coverage was similar among the three different water masses (STW, SAW, STF) (Table 1; Figure Suppl. S1 or S2).]

Samples were collected from 10 L Niskin bottles attached to a CTD rosette in association with a Seabird 9plus CTD, equipped with temperature, salinity, dissolved oxygen, fluorometer and beam transmissiometer sensors. During the TAN1516 voyage the CTD rosette was not available, so samples were collected with a Niskin bottle deployed using a messenger system manually down to 10 m depth and from the *R/V Tangaroa* Underway Flow-Through System (TUFTS) system equipped with temperature, salinity, and fluorescence sensors. Seawater samples for nutrients, chlorophyll a (Chl a) and DNA were sampled from the Niskin bottles using an acid-washed silicone tubing and filtered through different type of filters for processing.

* 1. Nutrients, total and size-fractionated chlorophyll a

Nutrient samples were filtered through Whatman GF/F filters into clean 250 ml polyethylene bottles and frozen at −20 °C until analysis using an Astoria Pacific API 300 microsegmented flow analyzer (Astoria‐Pacific, Clackamas, OR, United States) according to the colorimetric methods described in Law et al. (2011).

For total Chl *a*, 250-400 mL seawater were filtered using low vacuum (<200 mm Hg) through 25-mm GF/F filters. These were wrapped in aluminum foil and stored at -80 C until analysis. For size-fractioned Chl *a* (0.2-2-mm, 2-20-mm, >20-mm) 400-500 mL were filtered sequentially through 47-mm polycarbonate filters with 20-mm (by gravity), and 2.0-mm and 0.2-mm pore sizes (by vacuum). Filters were folded and stored in 1.5 mL cryovials at -80 °C until analysis (look into analysis methods of Hamilton lab).

* 1. DNA samples collection and extraction

Seawater samples of 1.5–5 L of were filtered either through 0.2 um filters (47-mm polyethersulfone, Pall-Gelman) using low vacuum or through 0.22 um Sterivex filter units (Millipore) using a peristaltic pump (Cole-Palmer). Disc filters were then folded and placed in cryovials and Sterivex units were filled with RNA Later and flash-frozen in liquid nitrogen prior to storing at -80 °C. Disc filters were cut in two halves first and then into small pieces using a sterile blade. Each half was placed in separate tubes and lysed in parallel (2 h at 65 C on Boekel thermomixer set at 750 rpm) using the Midi version of the Nucleospin Plant kit Midi Kit (Macherey-Nagel, Düren, Germany). The 100 uL of PL2 buffer recovered from each halved filter were then pool together and DNA extraction procedure continue with the Mini version of the Nucleospin Plant kit.

For Sterivex filters DNA was extracted using a Tris-buffered lysis solution containing EDTA, Triton X 100 and lysozyme (pH = 8.0) and the Qiagen DNA easy Blood and tissue. Briefly, RNA later was expelled into a 2 mL Eppendorf tube using a syringe and then centrifuge (13.000 rpm, 10 min). After removing the supernatant, the pellet was resuspended using 1 mL of the lysis solution and pipette back into the original sterivex. The cartridge was secured using parafilm, put into a 50 mL falcon tube and incubate in a shaking incubator overnight (75 rpm, 37 C). We then added 1 mL of buffer Qiagen buffer AL and 40 uL of proteinase K (20 mg/mL) into the sterivex. After securing the cartridge again we put it back into the Falcon tube and incubate for 2 hours (75 rpm, 56 C). After incubations the lysate was recovered from the cartridge and DNA extraction and purification continued following manufactures instructions in the *Qiagen DNA easy Blood and tissue* and *Purification of DNA from animal tissues* kits.

* 1. PCR amplification, amplicon sequencing and processing

The V4 region of the 18S rRNA gene was amplified using the eukaryotic primers V4F\_illum (5′ CCAGCASCYGCGGTAATTCC-3′) and V4R illum (5′-ACTTTCGTTCTTGATYRATGA-3′) with Illumina overhang adapters (Forward 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3′ and Reverse 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3′) and following procedures described in Gutierrez-Rodriguez et al. (2018). PCR reactions were prepared in 50 uL and using 2x KAPA HifFi HotStart ReadyMix, (0.3 nM dNTP, 2.5 mM MgCl2), 0.5 uM of each primer and 3 uL of DNA template (5-20 ng/L). The thermocycling profile include 95C/3min, 10 cycles (98C/10s, 44C/20s, 72C/15s), 15 cycles (98C/10s, 62C/20s, 72C/15s) and 72C/7min).

Amplicons sequencing was conducted at the Genotoul GeT core facility (Toulouse, France) using an Illumina Miseq and a 250 cycles Miseq kit v.2. The filtered 3.2. Summary of sequencing results. A total of 479 samples were sequenced generating a total of 9166190 reads with a median sequencing depth across all samples of 18954 reads per sample (range = 6539 – 36551) (Table 2) (Suppl. Fig. S3). Obtained sequences were processed using the DADA2 pipeline (Callaghan et al., 2016) following the procedure described in Trefault et al., (2021). Taxonomic assignation was done against PR2 data base version 4.12 (https ://pr2-datab ase.org/) yielding 20688 amplicon single variants (ASVs), 18882 of which were assigned to protist taxa. Details on the number of samples, reads and ASVs associated to each water mass are shown in Table 2.

* 1. Pre-processing of ASV table and diversity analysis (counts normalization, phyloseq, nMDS, ANOSIM, DESeq,…)

We standardized the ASV table to the sequencing depth in each sample by normalizing the relative abundance to the mean number of sequences obtained across samples (median sequencing depth \* (n\_reads\_ASV/total\_n\_reads\_sample)). The relative contribution of specific groups in different water masses and regions were estimated from the sum of standardized readsacross the samples considered.

Similarity analysis was done using non-multidimensional scaling (nMDS) and ANOSIM analysis using Phyloseq (McMurdie and Holmes, 2013) and Vegan (Oksanen et al., 2019) R packages. Differences in species abundance across waters masses and regions was analysed using negative binomial generalized linear models coded in DESeq2 package (Simon et al., 2010). For analysis of higher taxonomic rank (division, class) distribution and their relation to environmental variables, tax\_glom function in phyloseq was used to agglomerate previously standardized ASV table into the chosen taxonomic level.

FOR DISCUSSION

A closer look into TS variability reveals sharp gradients within this broad frontal zone and allow to distinguish between three different groups STW-influenced stations (STF-STW; SSS = 34.8 – 35.0 and T > 13 °C ), SAW influenced stations (STF-SAW; SSS = 34.4–34.6 and SST<13 °C), , and STF stations (STF-STF; SSS = 34.6 - 34-8 and T < 13 °C) consistent with previously reported observations in the STF zone over the Chatham Rise region (Zhou et al., 2012; Safi et al., in prep/submitted).

The number of DNA samples from STW (~200) were 2-fold higher than those from SAW and STF (~100 each). This difference was mainly due to the large number of samples from Spring Bloom II voyage (TAN1212) (Suppl. Fig. S1). Most of the DNA samples included in this study were taken from oligotrophic and mesotrophic waters (<0.5 mg Chla m-3) with only a few collected from waters with Chl *a* concentrations >1.0 mg Chla m-3. Seasonally, Spring and Summer were better covered than Autumn, while no samples from the winter months (June to August) were available.

3-RESULTS

3.1 Physical, chemical and biological variability between and within water masses

ST waters were identified as those with surface salinity >35 psu (range = 35.1 - 35.6) (Fig. 2) and included samples collected during the Cross-shelf (TAN1604), the Bay of Plenty (KAH1303), the Spring Bloom II (TAN1212), and several voyages that visited the northern mooring site (STM) of the Biophysical Moorings long-term program (Nodder et al., 2016) (Fig. 1, Table 1, Fig. S1). The Subtropical Front (STF) separating STW and SAW had salinity values ranging between 34.4 and 35.0 (Fig. 2) and included samples collected on the Chatham Rise during TAN1516 and Biophysical Mooring voyages as well as those collected in the STF upstream of the Chatham Rise, as it flows northwestward between the South Island of New Zealand and the Campbell Plateau (Fig. 1, Table 1, Fig. S1). SA waters presented salinity values <34.4 (Fig. 2) and included samples collected at the southern mooring site (SAM) located in the Bounty Trough, on Campbell Plateau, and the Subantarctic front (SAF) south of C. Plateau (Fig. 1, Table 1, Fig. S1).

Sea surface temperature was on average lowest in SAW (10.7 ± 2.4 °C, mean ± standard deviation, sd), intermediate in the STF (13.1 ± 1.7 °C) and highest in STW (16.1 ± 3.2 °C) (Fig. 2). Temperature showed greater overlap among water masses and regions than surface salinity (Fig. 2). ST waters sampled during the Spring Bloom II voyage, for instance, showed surface temperature consistently lower than 15 C (12.5 - 14.5°C) (Fig 2).

Nitrate concentrations were lowest in STW (2.98 ± 1.96 mmol/L), intermediate, and more variable in STF waters (4.28 ± 4.17 mmol/L) and highest in SAW (12.17 ± 4.02 mmol/L), consistent with HNLC conditions of these waters.

Chl *a* concentrations in the surface mixed-layer (Chl *a*ML) were on average higher in the STF (0.65 ± 0.65 mg Chl *a*/L) compared to STW (0.38 ± 0.31 mg/L) and SAW (0.37 ± 0.23 mg/L) (ANOVA, F(2,220) = 14.2, p< 0.0001). The smallest size-fraction (0.2 – 2.0-mm Chl *a*) dominated the phytoplankton community across all water masses but more so in STW (Chl *a* <2.0 ~75 % of total Chl *a*) compared to SAW and STF (40-50%, Fig. suppl. SFChla). The contribution of >20-mm size-fraction to surface mixed layer Chl *a* was higher in SAW and STF, and although it remained on average relatively low (<15%), it occasionally reached > 50% levels (Suppl. Fig. S4). [worth noting in the figure legend we only have SFChla data from TAN1516/Chatham Rise; TAN1702/Campbell Plateau, TAN1212/SpringBloom, TAN1203/SOAP; N = 102 surface mixed layer samples)

A closer look revealed regional differences in these physico-chemical and biological properties within each water mass. In SAW for instance, SAF surface waters were colder and fresher than those on the Campbell Plateau and in the Bounty Trough (Fig. 2). Surface nitrate concentrations were lower in the Bounty Through compared to Campbell Plateau and SAF, consistent with the southwards strengthening of HNLC conditions as expected. Chl aML concentration was higher on Campbell Plateau (0.62 ± 0.48, mean ± sd) compared to surface waters in the Bounty Trough (0.33 ± 0.20) and the SAF (0.21 ± 0.07) (Fig. 2), although differences were only significant between the Campbell Plateau and SAF south of it (One-way ANOVA, F(2,33) = 4.494, p = 0.019).

Within STW, surface temperature and salinity were highest in northernmost waters sampled during the Cross-shelf II voyage (TAN1604) and lowest in STW waters sampled during the Spring Bloom II voyage (TAN1212) conducted at the beginning of austral spring (September-October) (Fig. 2). Temperature and salinity at the Biophysical Mooring STM site were intermediate on average and had a greater range that reflected the wider temporal variability covered through multiples visits conducted at different times of the year (Table 1). Nitrate concentrations showed the opposite trend with highest values associated with the colder and deep-mixed waters, and lower values reflecting warmer and stratified waters of the Bay of Plenty and Cross-shelf II voyages (Fig. 2).

Regional differences were also observed between the STF flowing north of the Campbell Plateau, which transported colder and fresher waters, and STF further north flowing eastward over the Chatham Rise (STF, Chatham Rise) (Fig. 2). Nitrate concentrations tended to be higher in STF waters next to Campbell Plateau than on the Chatham Rise reflecting the HNLC nature of the plateau. Relatively high nitrate concentrations (>10 mmol/L) were also measured during the summer (TAN1516) on the Chatham Rise at some stations located in the southeastern flank of the STF with colder (10.7 and 11.3 °C) and fresher (34.47 and 34.56) characteristics of surface mixed-layer waters indicating subantarctic water influence (Fig. 1).

3.4. Alpha-diversity – Species richness

Species richness and the Shannon diversity index estimated in the euphotic zone were on average lower in the STF compared to SAW and STW (Fig. 3). Highest diversity in STW was observed in the northernmost waters surveyed during the Cross-shelf II voyage and at the STM site, which included samples collected during various cruises spanning multiple years and seasons. Protistan species richness during the Spring Bloom II was substantially lower than these other STW locations (Fig. 3). In SAW diversity was lower on Campbell Plateau compared to more oceanic waters of the SAF and the SAM site (Fig. 3). Within the STF, diversity was also lower in upstream waters of the STF located further south (e.g. Campbell Plateau) than downstream waters flowing over the Chatham Rise located at lower latitudes (Fig. 3). In this region, higher diversity values were associated with the STF site sampled during the voyages visiting the moorings possibly reflecting the greater temporal coverage of the Biophysical Mooring program (2009-12) (Table 1).

Species diversity tended to decrease with latitude (model I linear regression, F(1,478) = 72.65, R2 = 0.13, p-value < 0.0001, Suppl. Fig. S5), although differences in mean diversity values were observed among water masses and regions with similar latitudes (Suppl. Fig. S5). Similar relationships were observed between diversity and temperature (model I linear regression, F(1,208) = 74.75, R2 = 0.261, p-value < 0.001, Suppl. Fig. S5) with regional differences modulating this trend (Suppl. Fig S5). Within STW for instance, species richness was higher in oligotrophic waters and decreased with increasing Chl *a* being generally higher in warmer waters (Suppl. Fig. S6). Samples from the STF presented lower richness compared to SAW and STW across the entire range of Chl *a* and nitrate concentrations (Suppl. Fig. S6).

3.5. Beta-diversity of protist communities (nMDS, ADONIS statistical comparison)

Pending to run Redundancy or Correspondance Analysis

To explore the similarity between protistan communities we performed non-scaling multidimensional analysis using Bray-Curtis distances. A first analysis with all samples yielded two main clusters corresponding to samples from the euphotic and aphotic zone. (Stress = 0.14) (Suppl. Fig. S8). A second analysis focused on the euphotic zone (n = 210 samples) separated samples in groups corresponding to different water masses (Stress = 0.17) although certain overlap also occurred, particularly between some samples from the STF and SAW (Fig. 4). Samples from different regions tended to cluster separately as well, with the differences in the strength of this clustering between regions reflecting the spatial and temporal coverage.

To investigate the influence of different oceanographic drivers in the nMDS clustering we performed two PERMANOVA analysis one with several physical (T, Sal), chemical (NO3 concentration - euphotic zone median NO3) and biological (Chla concentration - water-column median) continuous variables, and a second one with water mass and region categorical variables that to a large extent integrates these oceanographic properties. In the first analysis, salinity explained 21% of the variability (F.Model = 31.4, P<0.001) followed by temperature (6.4%, F.Model = 9.5, P<0.001) while the nitrate and chla median concentration in the water column were also significant, but explained only 4% and 2% of the variability. Overall this set of parameters left 65% of the variability unexplained. The second PERMANOVA analysis showed that while the water mass explained 22% of the variability (F. Model = 21.0, P<0.001) – similar to that explained by salinity – the area/region explained an additional 17% of the variability (F.Model = 16.7, P<0.001) and up to 53% of the variability together with physico-chemical and biological continuous variables included in the first PERMANOVA analysis.

Pending to run Redundancy or Correspondance Analysis

NOTE: Indirect gradient analysis: Compare ranks obtained by nMDS with additional environmental parameters using spearman rank correlation. Envfit analysis performed with vegan packages using Hellinger transformed species and environmental tables yielded significant linear relationship between NMDS1 and NMDS2 axes and all environmental variables included in the analysis (T, Sal, oxygen, NO3, mixed layer depth, chla conc). However, the % variables explained (R2) was greater for T, Sal and Oxygen (0.5) than Nutrients, MLD and Chla (0.2-0.3)(See ANOSIM – ENVFIT results table.

Environmental interpretation of sample ordination – Constrained analysis/representation:

Direct gradient analysis: “Constrained (canonical) ordination analysis, only the variation in the species table that can be explained by the environmental variables is displayed and analyzed, and not all the variation in the species table” (Ramette et al., 2007).

In this study we also have some known gradients (T, Sal, N,…). So could explore species abundance or occurrence in response to those gradients using redundancy analysis or canonical correspondence analysis. Whether species abundance response to these gradients is the same in STW and SAW is something worth looking into. For instance, one would expect that nitrogen gradients to have different consequences for species distribution in STW (where nitrogen is considered the limiting nutrient) or SAW (where iron is considered the limiting nutrient).

In **RDA** the main axes are constrained to be linear combinations of the environmental variables. Two tables necessary one with species data and one with environmental variables. Multiple linear regression are used to explain variation between independent and dependent variables. Interest of this approach is: 1-to represent patterns of species variation as much as they can be explained by measured variables; 2- display the correlation between each species and each env. Variables in the data set. I’d like see I this approach can be applied to investigate relationship between specific phytoplankton classes (rather than species) and environmental variables. Produce a triplot with samples distribution along X % Y-axes, explained by env. Measurements; and arrows representing correlation of specific classes with each env. Variables. – See if we can adapt analysis from Faure et al., 2019 ISME J used with mixotrophic lineages. In **CCA** similar to RDA but based on unimodal species-environmental relationships rather than linear models. Explore shape of the distribution of species/class abundance along environmental parameters or gradients to see about linear or unimodal distributions. RDA and PCA linear-based methods more suitable for absolute abundance values. CCA used to model relative abundances and they also accommodate the presence of many zeros in the species table better than linear-based methods.

3.5. Taxonomic composition – Water masses

DIVISION/CLASS

* *Treemaps – using SUM of n\_reads*
* *Barplots mean (sd) most abundant division/classes*
* *Fig. 5A + suppl. S9A or 5B + suppl. S9B:*
* *Fig. 6A or 6B: Relative abundance box-plot water masses (or violin plots) : Remove this comparative description from the results and use it in the discussion section.*
* *Suppl. Fig. 6: Relative abundance – regional box-plot distribution – Same as above*

Across all samples, Dinoflagellata (34% of total - sequencing depth- normalized – reads) and Chlorophyta (27%) dominated the protist community (even after excluding Syndinales) in the euphotic zone. Ochrophyta (mainly diatoms and Pelagophyceae) and Haptophyta were the other most important phytoplankton taxa, while Radiolaria (4%), Marine Stramenopiles (MAST, 4%) and Ciliophora (3%) contributed most among the heterotrophic groups. Although such divisions were consistently dominant, their relative contributions and particularly, their composition at class and finer taxonomic resolution varied between water masses and regions (Fig. 5A/Suppl. Fig. S5 -merged?).

In STW for instance, where Dinoflagellate and Chlorophyta co-dominated the community (30% of sequences each), Mamiellophyceae accounted for the vast majority of reads affiliated to Chlorophyta (>95%) while Chloropicophyceae represented only a minor fraction of sequences affiliated to this division (Fig. 5A). Prymnesiophyceae, diatoms and Pelagophyceae (3-5% each) ranked next having similar mean relative abundance as heterotrophic MAST (5%), RAD-A (Radiolaria, 3%) and ciliates (Spirotrichea) (Fig. 5A). Cryptophycea (2%), Dictyochophyceae (1%) and Chrysophyceae (1%) represented a smaller fraction of the phytoplankton community in STW.

In SAW, the community was clearly dominated by Dinoflagellata/Dinophyceae (40% total reads). Followed by Chlorophyta (18%), Ochrophyta (15%) and Haptophyta (12%) with more even share of the total number of reads compared to STW (Fig. 5A). At class level, Chloropicophyceae (14%) rather than Mamiellophyceae was the most abundant group of green algae followed by Prymnesiophyceae (12%), diatoms (8%) and Pelagophyceae (6%) (Fig. 5A). The heterotrophic component in SAW was dominated by MAST (6%) followed by ciliates (3%) while the relative contribution of Radiolaria in the euphotic zone was minor (<0.5%) and mainly attributed to Polycystinea (Fig. 5A).

The composition of the protistan community in the broader STF presented intermediate characteristics between ST and SA waters (Fig. 5A). As in SAW, Dinoflagellates (39%) dominated the community although the contribution of Chlorophyta (26%) was on average higher and closer to levels found in STW. Accordingly, Mamiellophyceae was the second most abundant class (17%), but Chloropicophyceae (10%), and Prasino-Clade V (1.5%) made also substantial contributions. Ochrophyta (12%) accounted for a similar fraction of phytoplankton reads as in STW and SAW, but in STF was clearly dominated by Diatoms (10%) with Pelagophyceae (2%) and Dictyochophyceae (0.5%) representing a minor fractions of this division. The heterotrophic component was dominated by MAST (5%) and ciliates (4%) while the contribution of radiolarians (<0.2%) was on average lower than in STW and SAW (Fig. 5A).

GENUS/SPECIES

* *Figure 7 - Most abundant species – water masses- bar plot and treemap (Genus)*
* Suppl. Fig. S11 - Class specific – Genus/Species treemaps – water mass
* Figure 8. Heatmap with most abundant 50 species
* Figure 9 - DESeq STW vs SAW
* Suppl. Fig. S12 and S13: STW-STF; STF-SAW comparisons

The genus/species/ASV composition also varied among water masses (Fig. 7). In STW, *Ostreococcus lucimarinus* was the most abundant species in the euphotic zone followed by *Bathycoccus prasinos*. *O. lucimarinus* tended to dominate during the Spring Bloom II voyage and *B.* *prasinos* at the Bio-STM site, while they alternated in their dominance within the Bay of Plenty and Cross-shelf (Hauraki) regions (Fig. 8) These Mamiellophyceae species together with *Micromonas commoda* and other *Micromonas* species (*M. bravo I, II*, and *M. pusilla*) accounted for the majority of the sequences affiliated to green algae in STW (Fig. 7, Fig. 8 & Suppl. Fig. S11). Several dinoflagellates species such as *Gymnodinium\_*sp,, *Heterocapsa rotundata* and *Gyrodinium fusiforme* were among the most abundant species in STW. *Gymnodinium\_*sp, and *H. rotundata* were more abundant at the Bio-STM site whereas *G. fusiforme* thrived in the STF (Fig. 8).ASVs identified as *Polar-centric Mediophyceae\_X* sp. and *Minidiscus trioculatus* were the most abundant diatoms in STW, particularly in the Spring Bloom II voyage, while an unidentified *Pelagophyceae\_XXX\_*sp. and *Pelagomonas calceolata* mainly, accounted for most reads belonging to Pelagophyceae (Fig. 7, Fig. 8 & Suppl. Fig. S11). *Phaeocystis globosa* (ASV\_0065) was the most abundant species of Prymnesiophyceae (Fig. 7) with several ASVs belonging to *Chrysochromulina spp.* and *Gephyrocapsa oceanica* contributing to the overall dominance of these species among prymnesiophytes (Fig. 8, Suppl. Fig. S11).

In the STF, *Ostreococcus lucimarinus* was also the most abundant species overall, although the relative contribution of *Chloroparvula pacifica* (otu\_0014 - Chloropicophyceae) increased to levels similar to those attained by *Bathycoccus prasinos* (Fig. 7). It is worth noting the abundance of sequences affiliated to *Chloropicon sierburthii,* which togetherwith *C. pacifica* was responsible of the increased relative abundance of Chloropicophyceae in the STF (Fig. 7,, Fig. 8 and ). The heterotrophic dinoflagellate *Gyrodinium fusiforme* was the second most abundant species in STF with other ASVs affiliated with *Warnowia* sp. and *Karlodinium* sp. appearing among the most abundant dinoflagellate species (Fig. 7, Fig. 8). The higher relative abundance of diatoms observed in STF was mainly driven by *Fragilariopsis sublineata*, and *Thallasiossira* sp., which together with the same *Polar-centric-Mediophyceae* ASV found in STW (otu\_0030) were the most abundant diatom species in STF waters (Fig. 7 & Suppl. Fig S11). Among Prymnesiophyceae, *Phaeocystis* spp. was the dominant genus but most reads in this case belonged to *P. antarctica* instead of *P. globosa* (Fig. 7 & Suppl. Fig S11). The prymnesiophyte *Gephyrocapsa oceanica* also increased substantially in STF compared to STW waters. The contribution of Pelagophyceae in STF was relatively low (Fig. 6) and dominated by *Aureococcus* and *Pelagococcus* spp. (Fig. 8 & Suppl. Fig S11). The abundance of Cryptophyceae and Dictyophyceae remained minor overall (<2%) (Figs. 5), but both groups showed changes in their specific composition across water masses. Among Cryptophyta, *Plagioselmis prolonga* and *Teleaulax gracilis* increased substantially from STW to STF waters while the relative contribution of sequences assigned to *Teleaulax sp*. decreased. Changes within cryptophyta were milder but showed an increase in the relative abundance of *Dictyocha speculum* and *Pseudochattonella farcimen* from STW to STF waters (Suppl. Fig. S11)

In SAW, ASV\_0014 assigned to *Chloroparvula pacifica* was the most abundant taxa with other ASVs of this species (asv\_0086) contributing also to the dominance of Chloropicophyceae over Mamiellophyceae in these colder waters (Fig. 7, Fig. 8 & Supp. Fig. S11). Among the latter, *Bathycoccus prasinos* became the most abundant genus while *Ostreococcus lucimarinus* was virtually absent in SAW (Fig. 7, Fig. 8 & Supp. Fig. S11). The increase in the relative abundance of Prymnesiophyceae in SAW was driven mainly by *Phaeocystis* spp., with *Phaeocystis antarctica* emerging as the second most abundant species in SAW (Fig. 7), and other *Phaeocystis* species (*P. globosa*, *P. cordata* and *Phaeocystis* spp.) contributing also to the dominance of this genus among prymnesiophytes (Fig. 8, Supp. Fig. S11). The diatom *Fragilariopsis sublineata* (ASVs 0061 & 0036) was the dominant species of diatom in SAW (Fig. 7) while *Thallasiossira* sp., *Polar-centric-Mediophyceae* and other genera contributing substantially less (Fig. 8 & Supp. Fig. S11). The increased abundance of pelagophyte reads in SAW (Fig. 5) were mainly assigned to the *Pelagophyceae*\_XXX\_sp - the same ASV that dominated in STW - and to *Aureococcus anophagefferens* which was among the most abundant taxa in SAW (Fig. 7).

To investigate further differences in species composition between water masses we ran DESeq analysis using water mass as categorical variable. This yielded 109 and 62 ASVs out of 4572 ASVs (DESeq Suppl. Table) that were significantly more and less abundant (p-value < 0.01) in STW compared to SAW, respectively (Fig. 9a). Species that showed greatest differences (>10 log2-fold changes) were not necessarily among the most abundant species in each water mass. Among the species that showed greater affinity for STW we found the diatoms *Minutocellus polymorphus*, *Polar-centric-Mediophyceae* and *Minidiscus* *triculatus*; the prasinophytes *O. lucimarinus*, *C. sieburthii*, and various *Micromonas* species, the prymnesiophyte *Chrysochromulina* sp. and the dinoflagellate *Heterocapsa rotundata* (Fig. 9A). Among species with preferences for SAW we found several diatom species including *F. sublineata, Pseudo-nitzschia* sp.and *Cylindrotheca closterium*, the prymnesiophyte *Phaeocystis antarctica*, dinoflagellate species such as *Gonyaulax* sp. or *G. fusiforme,* and the pelagophyte *Pelagococcus* sp. (Fig 9A).

Different ASVs affiliated to the same species often showed very different (i.e., opposite) relative changes between water masses. Some ASVs identified as *Chloroparvula pacifica* that were highly abundant in SAW (e.g. otu\_0086 and out\_0014, Fig. 7) showed also greater affinity for SAW waters, while less abundant others (e.g. asv\_0532 and asv\_0336) showed significant preferences for STW (Suppl. Fig. S12). Most abundant ASVs of *P. antarctica* (asv\_0011, asv\_0027) showed preference for SAW (10-30-log2 fold negative change) while much less abundant asv\_0218 showed greater affinity for STW. Similar intraspecific variability was observed within more loosely defined ‘species’ (e.g. *Dinophyceae\_XXX\_*sp. *and Pelagophyceae\_XXX\_*sp.) which included ASVs with opposite affinities for STW and SAW (Fig. 9A). The broad spectrum in the magnitude and sign of the log2fold change between STW and SAW observed within ASVs affiliated within this and other ‘species’ highlight the ecological diversity encompassed by different strains (FOR DISCUSSION).

DESeq analysis between STF and STW yielded 46/72 (+/-) significant different abundances while only 25/24 (+/-) rendered significant differences between STF and SAW (Fig. 9B and 9C). In most cases, the distinctive species observed between the STF and either STW or SAW coincided with those identified from the comparison between STW and SAW described above (Fig. 9B and 9C). For instance, the diatom *Minutocellus polymorphus* found in higher abundance in STW compared to SAW, was also associated preferentially with STF compared to SAW while the diatoms *F.* *sublineata* and *Pseudo-nitzschia* sp., which showed significantly higher abundances in SAW compared to STW, were preferentially associated with the STF rather than STW (Fig. 9B and 9C). Only a few species such as *Thalassiothrix longissima* and *Chloropicon\_sieburthii* were distinctively associated with STF waters.

DISCUSSION

A major goal of this study was to characterize the diversity and taxonomic composition of protist community in SA and ST water masses and across the frontal zones that delimits them. Although samples included in this study were collected in different seasons, the seasonal coverage was similar for STW and SAW, and thus warrants a meaningful comparison between the communities associated with these water masses throughout the year. Previous phytoplankton studies in this region of the SW Pacific had been focused mainly on the STF over the Chatham Rise region and the SA and ST waters flanking this broad frontal zone (Chang et al., 1998, Hall et al., 2001, Delizzo et al., 2009). Using microscopy, pigment, and flow cytometry methods these studies described the prevalence of larger cells and diatoms through winter and spring in the more productive waters of the STF on the rise (Bradford-Grieve et al 1997; Chang et al., 1998). ST waters on the northern flank larger were also dominated by diatoms and nanoflagellates dominated in spring but dinoflagellates in winter, while smaller cells mainly composed by haptophytes, pelagophytes and dinoflagellates dominated the eukaryotic phytoplankton in SA waters (Chang et al., 1998; Bradford-Grieve et al., 1999; Peloquin et al., 2011). Here we build on this knowledge and greater taxonomic resolution and coverage of DNA metabarcoding to provide a comprehensive characterization of the diversity and taxonomic composition of protist communities associated to these contrasting water masses and new insights about ecological preferences of major taxa from class to species and intraspecific levels.

**Alpha diversity patterns**

Species richness was higher in ST than SA waters and lowest in the STF (Fig. 3). Richness decreased latitudinally and with temperature (Fig. S4) as expected from global diversity patterns (REF) and previous reports in the region (Raes et al., 2018). The lower diversity in the more productive STF was observed in both euphotic and aphotic layers (Fig. Suppl) on average lowest in more productive STF compared to SA and ST waters (Fig. 3, Fig. S5). This water-mass related pattern was systematically observed across the different levels of nitrate and chla concentrations encompassed in this study (Fig. 6). A study conducted across the Southland Current, a coastal expression of the STF that flows along New Zealand’s South Island, did not found a significant increase in protistan diversity across this coastal frontal zone (Allen et al 2020), while a decrease in eukaryotic richness across the STF in oceanic waters east of the Chatham Rise had been reported (Raes et al. 2018). These results together with the relatively lower diversity observed in the STF suggest a negative effect of increased productivity typically associated frontal zones upon protistan communities.

Regional differences in species richness were also evident within different water masses (Fig. 3). Most notably, species richness tended to be lowest in the Campbell Plateau region in both SAW and STF for which we have no clear explanation other than speculate about the possible effect of the hydrographic and bathymetric confinement and elevated productivity of this subantarctic region (Neill et al., 2004; Gutierrez-Rodriguez; Forcén-Vazquez et al., 2021). Within the other end of species richness represented by STW, the lowest richness was observed during the Spring bloom voyage certainly due to the Lagrangian sampling strategy of this voyage, but also to the expected decrease in diversity inherent to bloom conditions (REF). One could argue that more productive waters such as the STF, Spring Bloom, and the C. Plateau exhibited lowest diversity levels. Diversity patterns associated with water masses and regions are the same when considered only the surface mixed layer (Mixed-layer depth – 0.2 C change relative to 10 m) or the twilight zone (>150 m) (Supplementary Figure 1) suggesting that whatever the reasons behind the lower diversity associated with the STF it affected the entire water column.

Despite the regional and seasonal variability surveyed within STW and SAW water masses (Fig. 1, Table 1, Fig. 2) we observed systematic differences in their taxonomic composition (Fig. 4). Such water mass specificity has been observed in the prokaryotic (Galand et al., 2009 & 2010; Agogué et al., 2009; Seymour et al., 2012; Techtmann et al., 2015) and eukaryotic component (Hamilton et al. 2008; Raes et al., 2018) of microbial communities across different oceans. Among environmental drivers, salinity (21%) rather than temperature (4%) and nitrate concentration (2%) was the physico-chemical variable that explained best the clustering of samples we obtained with the nMDS (21% of the variability). Bray-Curtis dissimilarity of surface bacterial communities across the Southland current – the coastal expression of STF in that flows along the South Island of New Zealand - was also strongly correlated with salinity (Baltar et al., 2018). These results support the view that ST and SA water masses east of New Zealand are better conceptualized as bioregions or provinces rather than habitats (sensu Martiny et al., 2006), where (phyto)plankton communities reflect oceanographic processes and history in addition to contemporary physico-chemical conditions.

Samples from the STF itself were also distinguished from those in SA and ST water masses based on their taxonomic composition, although they showed a greater overlapping (Fig. 4) reflecting the active mixing and transition role of the STF. This overlap was particularly evident between samples collected at the Bio-STF and the Bio-SAM sites located on top of the rise and its southern SA flank and between the STF flowing north of the C. Plateau and the plateau itself, which suggests a stronger connectivity of the STF with SA than ST waters. Similarly, the horizontal mixing and phytoplankton community size structure in the STF zone has been reported to be more tightly coupled across the SA than the ST (Safi et al., submitted). This contrast with the higher similarity between diatom species of the STF and STW compared to SAW reported in this region (Chang et al., 1998). At any rate, the distinct communities observed in ST and SA, and to a lesser degree in the STF, highlights the role of oceanographic features such as the STF as boundaries that influence the diversity of oceanic microbial communities in large oceanic provinces (Baltar et al., 2016; Raes et al., 2018).

Samples from different regions tended to cluster separately as well, with differences in the strength of this clustering between regions likely reflecting differences in the spatio-temporal coverage of the sampling. We interpret that the relatively looser clustering of samples within each site of the Biophysical Moorings Time-Series (Bio-STM, Bio-STF and Bio-SAM) was due to the multiple season/years span of this program. Samples from the Spring Bloom II voyage were tightly clustered in a way consistent with the Lagrangian sampling strategy adopted in this voyage (and the limited compositional changes observed during the 20-days that the subtropical bloom was tracked) while samples from the Cross-Shelf voyage cluster more loosely reflecting the more diverse environments (coastal shelf, slope and oceanic waters) and communities surveyed during this voyage.

**Phytoplankton community composition**

Our results showed the overall dominance of dinoflagellates and chlorophytes followed by diatoms, haptophytes and pelagophytes (Fig. 5). Yet consistent differences in the taxonomic composition between water masses emerged at class and species levels belonging to these groups (Figs. 6 & 7). Analysis of intraspecific diversity revealed differences in the distribution of ASVs of the same species suggesting the presence of different ecotypes in some cases (e.g. Chloroparvula pacifica; Phaeocystis antarctica) and current taxonomic gaps within certain groups that remained to be characterized (e.g. Pelagophyceae\_XXX; Dinophyceae\_XXX)(Figs. DESeq). Below we discuss the distributional patterns of major groups zooming across different taxonomic ranks in an attempt to shed some light on their ecological preferences and the level of taxonomic resolution required to link the community structure and function.

In the case of green algae for instance, the relative contribution of the two main classes found, Mamiellophyceae and Chloropicophyceae showed opposite distribution patterns (Fig. 5). Mamiellophyceae was most abundant in STW and constituted the bulk of green algae dominating these waters while its relative abundance decreased in STF to reach lowest levels in SAW (Fig. 6). Its dominance in STW was mainly driven by *Ostreococcus lucimarinus*, with *Bathycoccus prasinos* and several species of *Micromonas spp.* (Fig. 7). *O. lucimarinus* was clearly the most abundant species here and in the STF (Fig. 7) in agreement with a metabarcoding analysis conducted across the Southland Current, coastal expression of the STF in the New Zealand South Island, where O. lucimarinus was identified as the most abundant species in the ST flank with decaying concentrations with increasing distance to coast and into SA waters (Allen et al., 2020). *O. lucimarinus* emerged also among the most abundant Mamiellophyceae species in a 18S rRNA metabarcoding survey conducted on coastal waters globally (Tragin and Vaulot 2019). In addition to *O. lucimarinus,* wealsoIn SAW the relative contribution of Mamiellophyceae was substantially lower (Fig. 6) and dominated by the species identified as *Bathycoccus prasinos* and *Micromonas commoda* with very low abundance of *Ostreococcus* (Fig. 7).

The dominance of picoplanktonic Mamiellophyceae is consistent with the greater contribution of <2 um chla (80%, Fig. S4) observed in STW. It is worth noting that the highest abundance of this group was observed during the onset of the spring bloom (TAN1212, Table 1) when Mamiellophyceae accounted for 40-75% of 18S rRNA reads and diatoms contribution remained around 10% over the 3-weeks sampling (Fig. Suppl. Regions bar plots). Prasinophytes have been reported to contribute substantially to vernal blooms in temperature latitudes (Bustillos-Guzmán et al., 1996; Latasa et al., 2010; Gutierrez-Rodríguez et al., 2011; Nunes et al., 2018). Mamiellophyceae and *O. lucimarinus* were the most abundant phytoplankton class and species in the ST flank of the STF zone (TAN1516, Fig. 1, Table 1) in agreement with high prasinoxanthine and prasinophyte relative contribution reported during spring in this region (Delizzo et al., 2009). High abundance of several species of prasinophytes including Ostreococcus and Micromonas have been recently reported during the onset of the North Atlantic spring bloom from 16S rRNA amplicon sequencing analysis (Bolaños et al., 2020). The deep mixing layers (>100 m) during the Spring bloom voyage (Chiswell et al., 2020) where they dominated supports their ability to thrive under high-nutrient high-mixing conditions and highlights their important role in the development of spring blooms characteristic of temperate latitudes. Overall, our results highlight the wide ecological breadth of Mamiellophyceae and certain species like *Ostreococcus lucimarinus* which tend to dominate across a wide range of physical, chemical and trophic conditions encountered within STW.

Chloropicophyceae showed the opposite trend with the highest relative contribution associated with SAW (Fig. 5, 6). Chloropicophyceae, previously clade VII, has been recently described as a new class of picophytoplanktonic prasinophytes (Lopes dos Santos et al., 2017a). Culture representatives of Chloropicophyceae have been obtained from tropical and subtropical latitudes of the north and south Pacific (Lopes dos Santos et al., 2017a and references there in) but to our knowledge this is the first report of their presence and high abundance in subantarctic waters. The majority of Chloropicophyceae reads were assigned to a reference sequence corresponding to *Chloroparvula pacifica* and included several ASVs one of which (ASV0014) was the most protist found in SAW (Fig. 7). Chloropicophyceae is considered to be the dominant group of green algae in oligotrophic oceanic waters in contrast with the preference of Mamiellophyceae for richer coastal environments (Lopes dos Santos et al., 2017b; Tragin et al., 2018). In this study, Chloropicophyceae were most abundant in SAW which in the region east of New Zealand are considered HNLC, suggesting that the preference of this group for oligotrophic conditions reported for typically macronutrient limited waters (REF) could also encompassed iron limited HNLC systems. It is noteworthy that highest abundance of this class within SAW were observed on C. Plateau, where biological evidence of natural iron fertilization has been reported (Gutiérrez-Rodríguez et al., 2020) and in the STF north of the plateau (Fig. suppl. Regions) where *Chloropicon sieburthii* made a substantial contribution in addition to the dominant *Chloroparvula pacifica*. Moreover, an ASV belonging to this genus was also found to contribute substantially to protistan community in coastal waters of the California Current Ecosystem (Gutiérrez-Rodríguez et al., 2018), highlighting the need of further studies to better understand the ecological drivers beyond coastal-oceanic trophic gradients responsible of the water mass preferences they showed. What do we know about Chloroparvula distributions and ecology?

The relative contribution of dinoflagellates tended to be higher in SAW and the STF compared to STW (Fig. 6), consistent with previous the microscopy-based observations (Chang et al., 1998). *Heterocapsa rotundata, Karlodinium spp., and Gymnodium spp*. were most abundant photosynthetic species in SAW and STF (Fig. 5). However, ASVs affiliated to the heterotrophic *Gyrodinium* genusand particularly *G. fusiforme* were identified as the most abundant species in agreement with previous study in the Southland current where in addition to *Karlodinium* and *Gymnodinium*, *Gyrodinium helveticum* and *G. spirale* were also retrieved among the most abundant species (Allen et al., 2020). This genus is known to be an important component of microzooplankton prevailing in high latitude coastal and oceanic waters (Archer et al., 1996; Strom et al., 2001; Olson et al., 2002) being active grazers with the capability of cropping down iron stimulated diatom blooms in the HNLC waters of the subarctic Pacific (Saito et al., 2005 & 2006). While species of *Gyrodinium* were prevalent across all water masses in our study, their abundance was higher in more productive STF waters, where higher chla concentration was accompanied by increasing abundance of diatoms and larger cells, confirming their pivotal importance in pelagic foodwebs as the link between phytoplankton and mesozooplankton (Zeldis et al., 2019).

As expected, diatoms tended to be more abundant in the STF compared to STW and SAW (Fig. 5) although relatively high contribution (>30%) were observed in all water masses at times (Fig. 6). Most abundant ASVs in the STF were identified as *polar-centric-Mediophyceae* (Fig. 7) with several large diatom species including *Fragilariopsis sublineata*, *F. cylindrus*, *Chaetocerus periuvianus*, *Cylindrotheca closterium* and *Thalassiothrix longissima* were significantly more abundant in STF compared to STW (Fig. DESeq analysis). In addition to Polar-centric-Mediophyceae, other small diatoms as *Minidiscus trioculatus* and *Minutocellus polymorphus* were most abundant diatoms in agreement with the dominance of these small diatom taxa in STW of the Southland Current (Allen et al 2020). While most common genus reported in STW (and STF) by microscopy analysis (e.g. *Thallasiosira spp., Chaetoceros spp., Guinardia spp*.) (Chang et al.,1998) were also detected by DNA metabarcoding analysis, small species revealed as numerically dominant by the latter approach tended to be overlook by microscopy. Diatoms are generally conceptualized as microplankton group associated with new production and high export potential (Legendre and Le Fevre, 1995; Vidussi et al., 2001; Uitz et al., 2006). However, there are increasing evidence supporting the importance of small nano or and even pico-size diatoms in both coastal and oceanic systems (Buck et al., 2008; Lomas et al., 2009; Leblanc et al., 2018; Hernández-Ruiz et al., 2018; Arsenieff et al., 2020). Our results showing the dominance of *M. trioculatus* and *M. polymorphus* in STW particularly during the more productive conditions of Spring Bloom II and Bio-NBM spring voyages (Fig. 8 heatmap or Fig. Suppl with regional most abundance species) further support and the important role of small diatoms in open-ocean phytoplankton proliferations.

In SAW, *F. sublineata* was clearly the most abundant diatom species (Fig. 6) with ASV0036 and ASV0061 being more abundant on C. Plateau and SAF waters, respectively (Fig. Suppl. regional most abundant species bar plot). The preference of *Fragiolariopsis* species for SA has been also reported by microscopy analysis (Chang et al., 1998). DESeq analysis revealed that 9 diatom ASVs were significantly more abundant in STF compared to STW, but only 3 of them (*T. longissima*, *Minutocellus polymorphus* and *Cerataulina pelagica*) were significantly more abundant in STF relative to SA. This suggests higher connectivity between the STF and SAW in agreement with tighter coupling between mixing and phytoplankton biomass size-structure reported in the southern flank of the STF (Safi et al., submitted).

Pelagophyceae showed the opposite trend compared to diatoms and have lowest contribution observed in STF waters (Fig. 6). The contribution of this class tended to increase with depth (Regional bar-plots show this – think about other way to illustrate their depth preference) in agreement with their preference for the DCM (Le Gall et al., 2008; Latasa et al., 2016; ….) and physiological adaptation to low light and high nutrient environments (Dimier et al., 2009’ Kulk et al., 2012; Dupont et al., 2015). This vertical segregation was evident in both STW and SAW despite the different specific composition observed in each water mass. *Pelagomonas calceolata* (ASV0044) was the most abundant species in STW while unidentified pelagophyte (Pelagophyceae\_XXX.sp, ASV0081) and *Aureococcus anophagefferens* (ASV0118) were most abundant species in SAW (Fig. 7). *Pelagomonas calceolata* is a widespread species (Andersen et al., 1993; Moon-van der Staay). Whether the ubiquity of this species is bound to high genetic diversity or physiological versatility is not clear although the low intraspecific diversity observed in pelagophytes assemblages globally (Cabello et al., PhD, submitted to EM) supports the latter. In our study, several ASVs were assigned to *Pelagomonas calceolata* and while the most abundant one showed preference for STW, other less abundant ASVs were significantly more abundant in SAW (DESeq figure). Similarly, we found different water mass preference among ASVs assigned to unidentified pelagophytes (Fig. DESeq). While these observations suggest that different ASVs may reflect ecologically relevant different units (Rodriguez et al., 2005; Farrant et al., 2016) they also highlight the importance of culture isolations and species characterization to better characterize the diversity of pelagophytes assemblages.

Prymnesiophyceae were prevalent across all water masses (Fig. 5) although their relative contribution tended to be lower (7% of total 18S reads, Fig. 6) than expected for open ocean waters. The prevalence of 19-Hex pigment marker in oceanic waters and the application of quantitative methods (e.g. CHEMTAX) have shown that prymnesiophyceae represents between 20-50 % of the phytoplankton community in oceanic waters (Andersen et al., 1996; Latasa et al., 2005; Liu et al., 2009; Swan et al., 2016). Such dominance has been also depicted by improved genomic approaches which revealed the extremely high genetic and functional diversity of non-calcifying prymnesiophytes (Liu et al., 2009; Cuvelier et al., 2010). In our study, non-calcifying species, mainly assigned to *Phaeocystis* spp. and *Chrysochromulina* spp., dominated the group (Suppl. Fig. Treemaps of specific groups). The abundance and relative contribution of *Phaeocystis* spp. was lowest in STW, intermediate in STF and peaked in SAW while Chrysochromulina spp. followed the opposite trend with higher contribution associated with STW (Fig. Treemaps of specific groups). The dominance of *Phaeocystis* spp. in SAW was mainly driven by *P. antarctica* (Fig. 7) in agreement with the prominence of this group in the SO (Verity et al., 2007) and observed decreasing abundance observed from SAW towards STW of the SW Pacific region during the Autumn-Winter season (Sow et al., 2020). Strains assigned to *P. globosa* and *P. cordata* were also detected in all water masses although they tended to be more prevalent and abundant in STW compared to SAW in agreement with previous reports in the region (Sow et al., 2020).

Coccolitophores are an important component of phytoplankton communities in the SO region extending from the STF to the Polar Front known as the Great Calcite Belt (Balch et al., 2011; Chang et al., 2016; Balch et al., 2016). In our study, higher abundances of coccolitophores were observed in the STF, in agreement with previous microscopy-based studies in this region of the SW Pacific (Saavedra-Pellitero et al., 2014; Rigual-Hernández et al., 2020). The most abundant coccolitophore species we found in our study were assigned to *Gephyrocapsa oceanica*, which contribute most in the STF but emerged among the 20 most abundant species also in SAW and present in STW (Fig. 7). However, *Emiliania huxleyii* that generally dominate coccolitophore assemblages in this region (Saavedra-Pellitero et al., 2014; Chang et al 2016) and in the SO (Holligan et al., 2010; Balch et al., 2016) showed very low abundances (data not shown).

*Cryptophyceae* contribution was relatively low on average (<3%) but showed lower abundance in SAW compared to STW where they represented up to 10% of protistan reads in the euphotic zone (Fig. 6). Cryptophytes are an ubiquitous phytoplankton group with widespread distribution from coastal to ocean systems and tropical to polar latitudes (Buma et al., 1992; Piwosz et al., 2013; Nunes et al., 2018). They have been reported to form blooms in coastal embayments (Jeong et al., 2013; Johnson et al., 2013) and coastal antarctic waters favored by low salinity conditions (Moline et al., 2004; Schofield et al., 2017; Nunes et al., 2019). The higher contribution we observed in STW relative to SAW in this study argue against the direct influence of salinity at least in open-ocean waters. The genus and species composition of the group also differ substantially between ST and SA waters (Fig. S11). In STW, *Teleaulax spp*. (ASV0107) was the dominant species followed by *Plagioselmis prolonga* (ASV0102) and *Geminifera cryophile* (ASV0192), while it was virtually absent in SAW where the same strains of *P. prolonga* and *G. cryophile* became the dominant cryptophyte species, although at much lower abundances than observed in STW. Cryptophytes have been observed to response positively to iron fertilization in HNLC waters of the North Pacific Sato et al., 2009; Suzuki et al., 2009) suggesting that their lower abundance in SAW in this study could be related to iron limited conditions characteristic of the region. The contribution of cryptophytes in STW was highest during the open-ocean spring bloom (TAN1212) and in shelf-slope stations (TAN1604) consistent with their preference for richer conditions (Fuller et al., 2006; Latasa et al., 2010; Carreto et al., 2016 PO). Significant contribution of cryptophytes has been also observed in open ocean waters of the NW Mediterranean at the termination of the spring bloom (Vidussi et al., 2000) where they even dominated the surface mixed layer community in highly stratified stations (Latasa et al., submitted PO). Interestingly, the higher contribution of cryptophytes in our study occurred towards the end of the Spring bloom surveyed (TAN1212)(Fig. Suppl. Regional Bar plots), coincident with strong surface stratification and biomass accumulation (Chiswell et al., 2019) supporting the greater importance that stratification may have on the group compared to salinity itself (Latasa et al., submitted).

**Heterotrophic groups composition and vertical distribution / Aphotic community composition from biophysical moorings time series**

To be completed.

DISCUSSION STRUCTURE OPTIONS

1. **For Scientific Reports:**
2. Light discussion structured in:
3. Alpha diversity differences between water masses and patterns in relation to environmental factors.
4. Phytoplankton groups composition (class, species, intraspecies) and distributions in relation to environmental factors.
5. Heterotrophic groups composition and vertical distribution / Aphotic community composition from biophysical moorings time series
6. **For other more ecological/oceanographic journals Alternative thematic structure ( DSR, EM, NZMSS for instance)**
7. Alpha Diversity and species patterns
8. Protist groups composition in major water masses
9. Protist groups composition across major frontal zones STF and SAF
10. Intraspecific diversity – Ecotypes (collide results about different ASVs from different groups in this section)

XXXXXXXXXXXXxXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXxXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXxXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

DISCUSSION

Study spatial-temporal coverage – Importance of regional variability

* *Implications of regional/temporal variability for interpretation of water mass preferences*
* *Discuss temporal/seasonal variability within each water mass with focus in biophysical mooring voyages that include Feb, March, May, Sept, Oct.*

Richness patterns

* *Low diversity in STF despite heterogeneity of conditions?*
* *Lower diversity in productive systems?*
* *Low diversity in C. Plateau?*
* *Rarefaction curves – diversity saturation? – sampling effort?*

One could argue that more productive waters such as the STF, Spring Bloom, and the C. Plateau exhibited lowest diversity levels. [**NOTE**: Diversity patterns associated with water masses and regions are the same when considered only he surface mixed layer (Mixed-layer depth – 0.2 C change relative to 10 m). However, differences in diversity between water masses was smoothed when only samples from the twilight zone (>150 m) where considered (Supplementary Figure 1).]

Community composition and environmental drivers

* *Class-level group distribution in relation to temperature, salinity*
* *Class-level group distributionin relation to trophic status of the system (Chla concentration or index with POC,…similar to Latasa et al., 2010)*
* *Groups distribution in relation to SFChla – e.g. diatoms pico and nano vs micro. How does the relative abundance of this group vary with community size structure – does the diatom-large phytoplankton general association sustained in our dataset (e.g. Minutucellus or Minidiscus contribution to diatoms signal does not imply an increase in community size structure)*

Here, their abundance tended to increase with temperature indicating this group preference for STW influenced waters (Figure 8 – suppl?). In STW though, they peaked in relatively cold (and deeply mixed) waters during the Spring Bloom II but decreased in warmer waters further north (Figure 8 – suppl?).

Euphotic vs aphotic – Biophysical moorings

Regional variability tax composition rel abundance of groups

Influence of regional and temporal variability on differences between water masses for discussion

* Fig. % abundance – Box-plot regional
* Species abundance – Regional
* Fig. % abundance – Env. Variables

Both alpha and beta diversity varied regionally (and temporally) within the three main water masses explored. In STW for instance, Chlorophyta dominated the community during the Spring Bloom II, driven mainly by the extremely high abundance of Ostreococcus (Mamiellophyceae) which accounted for approximately 50% of total reads (Figure treemaps – regional). The contribution of dinoflagellates during Spring Bloom, was substantial, yet minimum (<20%) compared to other STW regions (Figure Box-Plot regional). In well-stratified oligotrophic ST waters surveyed during late autumn (TAN1604), dinoflagellates were the dominant group (>40% total reads) followed by Chlorophyta (15%), with heterotrophic and mixotrophic MAST and Radiolaria increasing their contribution to similar levels (15%) (Suppl. Fig. treemaps – regional). Pelagophyceae tended to increase their contribution compared to the Spring Bloom II and Bio-NBM regions, while Prymnesiophyceae showed the opposite trend (Figure Box-Plot regional). Most abundant species,…need to re-do the plots with Bay of Plenty - Areas

The taxonomic composition of communities associated with different regions within SAW was also different. The Bio-SBM site located in the Bounty Trough south of the Chatham Rise was clearly dominated by dinoflagellates (40% reads) followed by Chlorophyta (20%) which included similar proportion of reads affiliated to Chloropicophyceae and Mamiellophyceae. Prymnesiophyceae and Pelagophyceae ranked next and accounted for a larger proportion of reads compared to diatoms (Figure regional - treemaps). The community on C. Plateau showed a larger contribution from Chlorophyta, which was dominated by Chloropicophyceae with a minor contribution from Mamiellophyceae. The contribution of MAST and diatoms increased while that of Prymnesiophyceae, Pelagophyceae, Dictyophyceae and Dinophyceae was minimum compared to SAW to the north and south of the C. Plateau. The relative contribution of diatoms increased further in colder waters of the SAF to the south together with Prymnesiophyceae they became the dominant phytoplankton/photosynthetic groups. Chloropicophyceae remained the most important group within Chlorophyta, although the contribution of green algae was lowest among SAW regions.

Similar changes at class level were also observed between the STF flowing north of the C. Plateau where Chloropicophyceae (& Prasino Clade V) were most abundant classes of green algae, and the STF over the Chatham Rise, where Mamiellophyceae became clearly the dominant class (Figure regional – treemaps).

The diversity of communities associated with the STF flowing north of the C. Plateau and over the Chatham Rise further north-east was also different (Figure – treemaps – regional – Pending to re-do the plots

STW surveyed in this study span several latitudinal degrees (33 and 41 S), seasons (spring, summer, autumn) (Table 1) and water-column conditions ranging from strongly-stratified oligotrophic waters (e.g. Cross-shelf voyage) to deep-mixing mesotrophic conditions (Spring Bloom II) and in-between conditions encounter during multiple voyages to the Bio-NBM site north of the Chatham rise (Map/Table 1).

**[5] Additional introduction/discussion points**

* Trophic index – classification similar to Latasa et al., 2010 about trophic preferences of phytoplankton chemotaxonomic groups but using taxonomic classes from DNA metaB?
* Functional groups – Taxonomic resolution with ecological significance – inferred from different distributional patterns and relation to physico-chemical drivers (T, Sal, Nutrients, …) division/class/genus/species What’s the taxonomic rank at which we see differences across water masses?
* Intraspecific variability – Differences in the spatial variability of ASVs of same species and implications for the interpretation of their functional groups and trait-based approaches. E.g. Diatoms (pigment-based) linked with accumulation of larger phytoplankton and export potential. Well, not always, if diatom increase is driven by small diatom species – Minutocellus, Minidiscus,…

NOTES on FIGURES from SN:

Figure 1 – need a diagram that shows the circulation and water masses

Suppl Figure S1 – seasonal differences across voyages? In S2?

Figure 3: need to think what these estimates of alpha diversity really mean…..

Suppl Figures S6 and S7– not sure what the Ranks refer to in terms of concentrations…..

Figure 6A and B – how are these abundances standardized? Why do they look like they do?!

FYI Cross-shelf and Spring Bloom II figures - Depth trends look interesting but need to be able to reference the samples to the depth