How The Environmental RNA Signature of Nitrate Reductase Can Be Used as a Predictor for Phytoplankton Distribution

Amanda Gardiner

University of Washington

School of Oceanography, Box 357940

Seattle WA 98195-7940

amandamg@uw.edu

Abstract

With anthropogenic climate change altering the ocean, it is critical to understand what variables control phytoplankton distribution so we can predict how distributions will be altered as result of climate change. This is vital because phytoplankton are important for primary productivity, as well as acting as the base of the marine food web supporting fisheries. This study aimed to determine whether the environmental RNA signature of Nitrate Reductase (NR) from phytoplankton is correlated with in-situ nitrate concentration. First, a phylogeny of phytoplankton NR amino acid sequences was generated to determine if the sequences separate into monophyletic groups matching their taxonomic identification. Secondly, RNA transcripts for NR derived from samples collected in the North Pacific Ocean were placed on the tree to determine which clades expressed NR. Finally, heat maps were constructed to show RNA abundance for each phytoplankton clade by latitude and nitrate concentration to establish distribution trends by phylum. All phytoplankton, no matter phyla or size, were more abundant in higher nitrate concentrations. When RNA abundance was normalized by chlorophyll concentrations, there was little separation in distribution based on plankton size, and different trends based on phyla emerged. Alveolata, Archaeplastida, and Stramenopiles were found in similar low to moderate nitrate concentrations (0.0023-0.8052μM). Cryptista did not display a consistent trend across the phylum, as all clades displayed different abundance patterns. Haptophyta, large and small, made up a significantly higher proportion of phytoplankton found in low nitrate environments (0.0009μM). These results indicate there is a separation of phytoplankton phyla by nitrate concentration, supporting the hypothesis that these phyla have evolved to utilize different ecological niches, however further research is needed with higher

taxonomic resolution to fully quantify the factors controlling the distribution of different phytoplankton clades.

Plain Language Summary

The goal of this study was to determine whether nitrogen availability in the ocean controls the distribution of different phytoplankton species. This was done by looking at the correlation between the abundance of a phytoplankton protein important for nitrogen chemistry, Nitrate Reductase, and the concentration of a specific nitrogen species, nitrate, in the ocean. All data used in this study were from research cruises in the North Pacific Ocean and publicly accessible protein databases. First, protein information was used to establish the relationships of phytoplankton to each other based on their Nitrate Reductase proteins. Then, RNA coding for Nitrate Reductase was matched to the different phytoplankton to see which groups of phytoplankton were present in the North Pacific, and their relative abundances. Finally, the RNA and nitrate concentrations were plotted by latitude to show what phytoplankton types were present at different nitrate concentrations. The first result was that phytoplankton of all sizes and types were found in higher abundance when nitrate concentrations were higher. When the RNA data was normalized according to phytoplankton abundance, the phytoplankton groups displayed different distributions. Three groups were found in similar low to moderate nitrate concentrations, one group was found in extremely low nitrate concentrations, and another group had no clear pattern based on nitrate, but instead a different pattern for each type of phytoplankton within the group. This indicates that different phytoplankton groups are found in

separate locations based on nitrate concentrations, and more research should be done to see what other variables control phytoplankton location.

Introduction

Despite only representing 0.2% of the global primary producer biomass, phytoplankton are a vital part of the biosphere (Field et al., 1998). They are responsible for creating approximately half of global net primary production, due to having a turnover time 1000 times faster than terrestrial plants (Field et al., 1998). In addition, phytoplankton serve as the primary producers at the base of the marine food web, so when phytoplankton communities change distribution or composition it can have a widespread effect on organisms of all trophic levels throughout the oceans (Brown et al., 2010). This is a result of the vast majority of the marine food web depending on phytoplankton. If phytoplankton abundance decreases, there is less food for herbivorous consumers, causing them to decrease in abundance, and therefore remove food sources for all other higher trophic levels (Fig. 1).

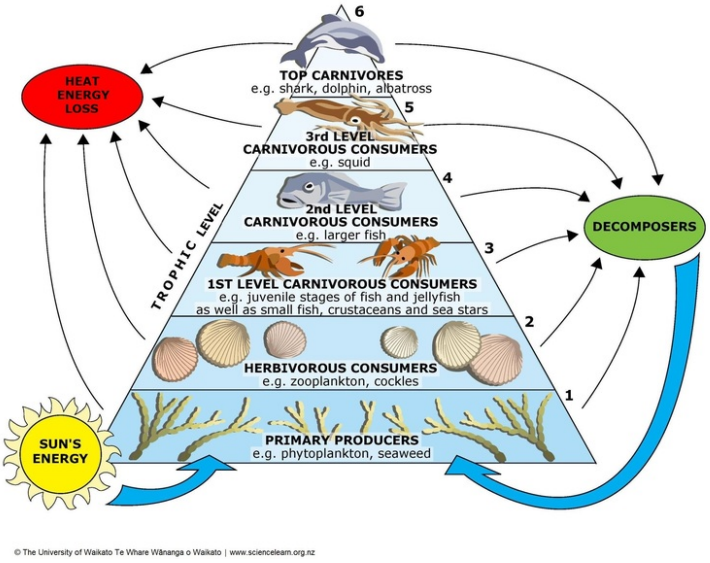

Figure 1. Marine trophic levels with example organisms in each level. Credit: The University of Waikato Te Whare Wānanga o Waikato

The importance of phytoplankton has led to extensive research attempting to quantify the number of species in the ocean, which has shown that they are incredibly diverse. Falkowski et

al. (2004) showed that in comparison to land plants, which are dominated by the single clade Embryophyta, phytoplankton species are distributed amongst eight major phyla (Fig. 2). This diversity has made it difficult to study the distribution of phytoplankton throughout the oceans, since the distributions of different species are affected by a multitude of factors (Harrison et al., 2010). For example, research in the NE Atlantic and Arctic Oceans has shown that phytoplankton species occupy unique ecological niches, with some species thriving in low temperature and high salinity water, while others thrive in warmer temperature and lower salinity water (Degerlund and Eilersten, 2010).

Research off the coast of Northern California has shown similar trends, with diatoms dominating in high nitrate coastal waters, whereas picophytoplankton are dominant farther offshore in lower nutrient oceanic waters (Chavez et al., 1991). Much of the research on phytoplankton distribution and primary productivity by phytoplankton has demonstrated a strong correlation between nitrate concentrations along with other nutrients that affect nitrate uptake like iron (Harrison et al., 1999; Martin and Gordon, 1988).

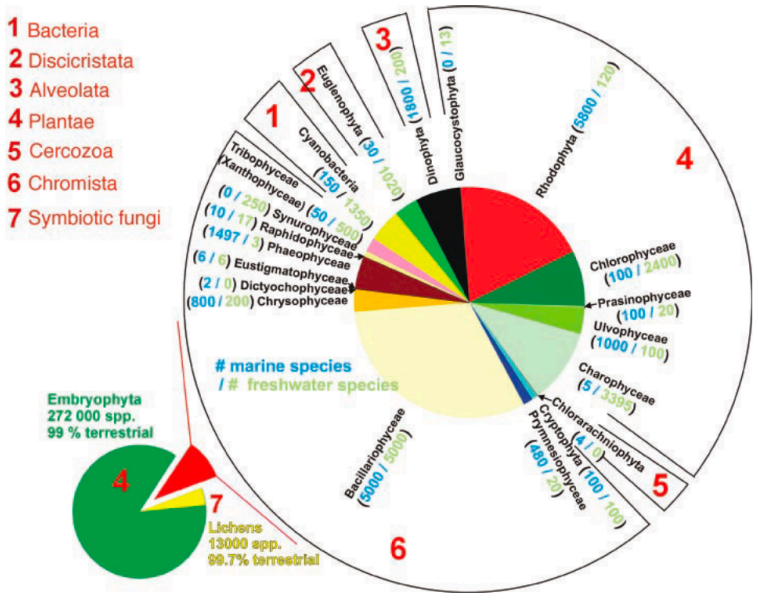

Figure 2. Phylogenetic distribution of both terrestrial and aquatic primary producers, which showcases how diverse phytoplankton are in comparison to terrestrial plants. The red numbers correlate to each major phytoplankton group, while the blue and green numbers are the number of marine and freshwater species, respectively. From Falkowski et al., 2004.

Historically, researchers have studied phytoplankton distributions through collecting environmental data on oceanic cruises and comparing the concentrations of nutrients essential for phytoplankton survival to the concentration of chlorophyll at that same station, to determine if there is a correlation between the two. A good example of this method is research by Martin and Gordon (1988), who showed that iron limits phytoplankton growth and is required for nitrate uptake. Chlorophyll is found in all phytoplankton species since it is required for light absorption and photosynthesis (Baldauf, 2003), thus chlorophyll concentrations indicates the relative biomass of phytoplankton present at a given location. The biomass can be used to estimate what size of phytoplankton are present, as large phytoplankton like diatoms dominate in high chlorophyll concentration areas, while small plankton like picophytoplankton dominate in areas with low chlorophyll concentrations (Harrison et al., 1999). However, chlorophyll concentrations do not give enough resolution to distinguish between different phytoplankton species, instead providing the broad trends of where phytoplankton are found and the likely dominance of small or large phytoplankton in a given area.

Recent studies, like that done by Simmons et al. (2016), examine genetic sequences from ocean water samples to identify which phytoplankton species co-occur in similar environmental conditions. This genetic data provides a vital opportunity, because studies can focus on the genes coding for important proteins involved in processes key for phytoplankton survival. Doing this can show why certain phytoplankton species thrive in different conditions, due to the fact that a phytoplankton's genotype will control the phenotypes of their proteins, which can affect protein activity.

An important enzyme that can be studied using this method is Nitrate Reductase (NR), which catalyzes the rate limiting step of nitrate uptake (Campbell, 1999). Nitrate Reductase is vital protein since nitrate is a key nutrient for phytoplankton (Howarth and Marino, 2006). Nitrate uptake by the enzyme Nitrate Reductase in phytoplankton follows Michaelis-Menten kinetics, which means that until the enzyme is saturated, the rate of nitrate uptake will increase as the concentration of nitrate in the environment increases (Eppley and Thomas, 1969). Due to the fact that natural selection will select for different rates of nitrate uptake in different regions, different gene sequences coding for Nitrate Reductase will be selected for in different regions, and therefore phytoplankton species will be found in different regions based on environmental nitrate concentrations.

Understanding the physiological reason behind phytoplankton thriving in different environmental conditions is important because it can be used in the prediction of how the composition of phytoplankton communities and their distributions will be affected by climate change. As climate change continues to warm the ocean, ocean stratification will increase, leading to decreased nutrient availability for phytoplankton (Steinacher et al., 2010). This can cause phytoplankton communities to change in composition, and be dominated by species that can thrive in the new conditions. For example, large phytoplankton like diatoms thrive in high nutrient environments due to having a low surface area to volume ratio which limits the rate of nutrient uptake, while small picophytoplankton are favored in low nutrient environments due to maximizing their rate of nutrient uptake by having a high surface area to volume ratio (Litchman and Klausmeier, 2008). This means that picophytoplankton will dominate in areas where anthropogenic caused stratification has increased and nutrient levels decreased, while diatoms

and other large plankton will shift to higher latitudes and cooler water with a deeper mixed layer (Barton et al., 2016). This alteration in phytoplankton distribution will cause phytoplankton biomass to decrease in certain areas such as around Hawaii (Fig. 3) where fisheries rely on phytoplankton as a food source (Howell et al., 2013). This will cause large scale changes in fisheries that people rely both for food and income, with many fisheries moving to higher latitudes, and with catches overall decreasing (Cheung et al., 2009). While the degree of change varies by ocean, one of the places that is predicted to have the largest decrease in global Net Primary Productivity, and as a side effect have drastic changes in fisheries productivity, is the North Pacific Ocean (Cheung et al., 2009). Specifically, one model predicts a 10-22% decrease in phytoplankton biomass by the end of the 21st century around Hawaii (Howell et al., 2013). With climate change threatening global marine ecosystems, and the lives of many people, it is imperative to implement policy changes to minimize major shifts in phytoplankton populations. This cannot be done until the distributions of phytoplankton throughout the oceans are accurately quantified. This study aims to determine whether the genetic signature of Nitrate

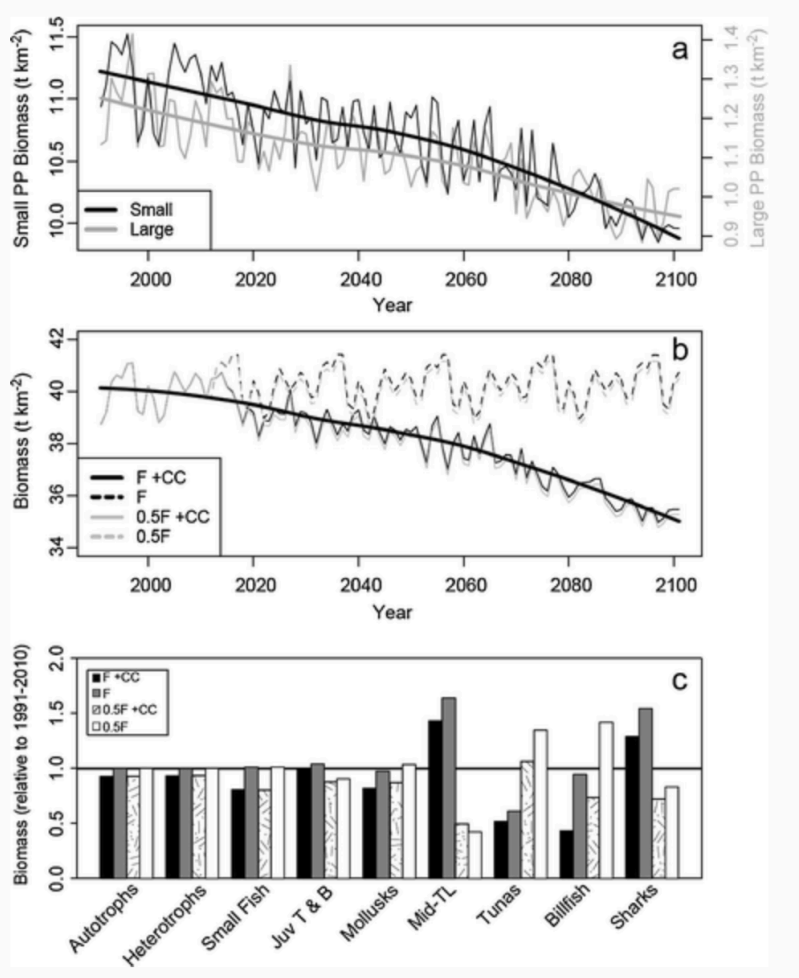

Figure 3. Climate Model predictions for changes in phytoplankton biomass (a), total marine biomass (b), and biomass per trophic group (c). CC stands for climate change, and F stands for fishing at current rate. From Howell et al., 2013.

Reductase in phytoplankton is strongly enough correlated to the environmental nitrate concentration to be used as a predictor of phytoplankton location.

Materials and Methods

All data used in this study was either from open access databases such as the NCBI or from data gathered on the Gradients 1 and Gradients 2 cruises (Fig. 4). Both cruises have their data and metadata available on SimonsCMAP (https://simonscmap.com/).

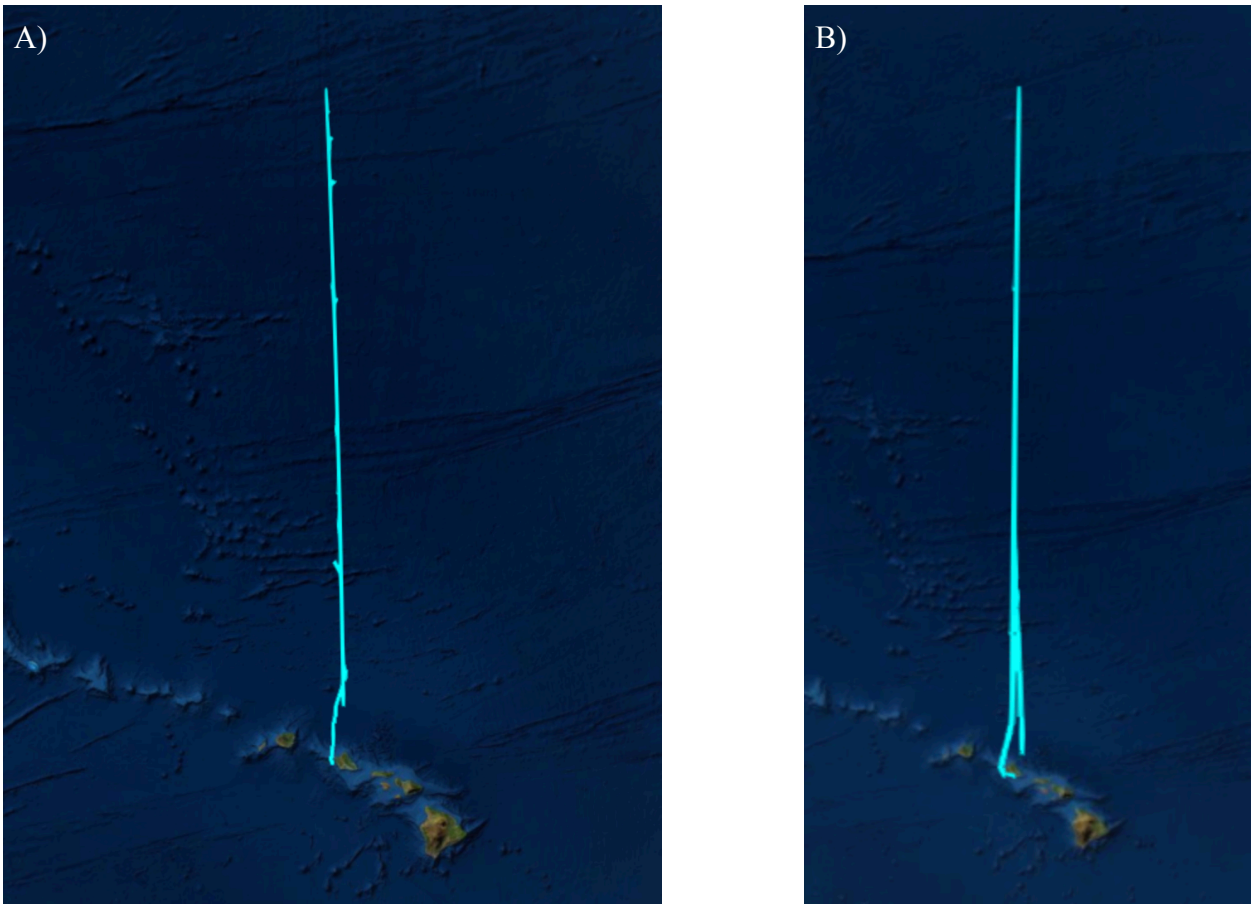

Figure 4. Visualizations of the routes taken by (A) Gradients cruise 1 and (B) Gradients cruise 2, which started and ended in Hawaii. Visualizations are from SimonsCMAP (https://simonscmap.com/). Gradients 1 started at 21.4542ºN and ended at 37.8864ºN. Gradients 2 started at 21.3008ºN and ended at 42.427ºN

Nitrate Reduce Reference Phylogeny

A reference phylogeny was constructed to compare the similarity in amino acid sequence of the Nitrate Reductase (NR) enzyme between different phytoplankton species. The accession numbers in the supplementary information from Wang et al. (2018) for NR sequences of 10 phytoplankton species (Table 1) were obtained, and the sequences downloaded from the NCBI database. The 10 sequences were aligned using MAFFT software version 7.475 (https://mafft.cbrc.jp/alignment/software/) and then using Hmmer software version 3.3.2 (hmmer.org), the 10 sequences were used to create a Hmm profile, which was run through the Marine Reference Database II to find homologous sequences in other phytoplankton species. 22,044 amino acid sequences mapped to the Hmm profile.

Table 1. Accession numbers for Nitrate Reductase amino acids sequences of the 10 different phytoplankton species used to create the Hmm profile. Species and accession numbers found in supplementary information from Wang et al., 2018.

Species	Accession Number
Chlamydomonas reinhardtii	AAF17595.1
Chondrus crispus	CDF 34670.1
Cylindrotheca fusiformis	AAY59538.1
Dunaliella salina	CBX85149.1
Emiliania huxleyi	DAA12507.1
Gracilaria tenuistipitata	ACX31652.1
Heterosigma akashiwo	AER70124.1
Klebsormidium nitens	GAQ91960.1
Phaeodactylum tricornutum	AAV66996.1
Thalassiosira pseudonana	EED88244.1

Table 2. List of species with NR3 isoform like gaps in their NR sequences that were added to the NR reference phylogeny, and the clade they belong to

Species	Clade
Alexadrium monilatum	Dinoflagellate
Azadinium spinosum	Dinoflagellate
Heterocapsa rotundata	Dinoflagellate
Karenia Brevis	Dinoflagellate
Polarella glacialis	Dinoflagellate
Pseudo nitzschia fraudulenta	Pennate Diatom
Pyrodinium bahamense	Dinoflagellate
Scrippsiella trochoidea	Dinoflagellate
Togula jolla	Dinoflagellate

Using seqmagick software v0.8.4 (https://github.com/fhcrc/seqmagick), the 22,044 sequences were sorted in order from most homologous to least homologous, and then the most homologous strain for each phytoplankton species was chosen for the reference tree. Sequences from the first 51 species were taken, and then an additional nine sequences were added (all species were in the top 200 most homologous strains) because they had gaps in their aligned sequences similar to those observed in the NR3 isoform of NR observed by Wang et al. (2018). NR sequences from 72 species in total were used to build the reference tree: the 10 sequences used in the Hmm profile, the top 51 species with most homologous sequences, 9 species with NR3-like gaps in their sequences (Table 2), and then the NR2 and NR3 strains of *C. subsalsa* from Wang et al. (2018) to compare with the species found with similar NR3 like gaps. Seqmagick was used to align all 72 sequences, and then Jalview V2.11.1.3 (https://www.jalview.org) and Trimal V1.2 (http://trimal.cgenomics.org) were used to trim the ends and major gaps from the sequences. RAxML V8 (Stamatakis, 2014) was used with the maximum

likelihood method to create a molecular phylogeny using the 72 trimmed and aligned sequences, and bootstrap was run 100 times for support. Nodes on the tree with bootstrap values less than 10% were collapsed into polytomies. The resulting phylogeny was visualized and annotated in iTol (https://itol.embl.de).

Environmental RNA Placement onto Reference Phylogeny

The next step was to create a 'fat plot' by placing environmental RNA data taken from the Gradients 1 and 2 cruises onto the NR reference phylogeny. The 72 aligned and trimmed sequences were used to create a Hmm profile with Hmmer software version 3.3.2 (hmmer.org). This Hmm profile was run through the environmental RNA transcript data from Gradients 1 and 2 cruises to find homologous RNA sequences that would generate amino acid sequences similar to those found in the tree. The hmm-align function in Hmmer software was used to align the environmental RNA sequences to the Hmm profile, and then Pplacer V1.1 (https://matsen.fhcrc.org/pplacer/) was used to place all the environmental RNA sequences onto the reference tree by which sequence they were most homologous to. The resulting tree was visualized and annotated in iTol.

Heat Maps

The data used in the heat maps was the RNA environmental transcript data and nitrate plus nitrite concentrations sampled during the Gradients 1 and 2 cruises, in addition to their meta data and a taxonomic identification file provided by the Armbrust Lab. The nitrate and nitrite concentrations were sampled three times between 0m and 15m depth, and were averaged to get

the mean concentration across the mixed layer. To create the heat maps, R (https://www.R-project.org/) was used with R studio (http://www.rstudio.com/) and the packages viridis (https://CRAN.R-project.org/package=viridis), dplyr (https://CRAN.R-project.org/package=dplyr), gplots (https://CRAN.R-project.org/package=gplots), reshape2 (http://www.jstatsoft.org/v21/i12/), and ggplot2 (Wickham, 2016). The number of environmental RNA transcripts per clade was plotted by latitude for each clade, and separated by cell size; based on filter size used in sample processing. The two filter sizes used were 0.2µm and 3µm, which would collect cells 0.2-3µm in size and >3µm in size respectively. Heat maps were plotted with raw transcript/L data from each clade at each latitude or had the transcript/L data normalized with respect to chlorophyll concentrations at each station sampled to account for bias based on abundance. All heat maps have z-score normalization, which means the scale shows how many standard deviations the RNA abundance is away from mean. All clades plotted on the heat maps were organized by what phylum they were in to make it easier to discern broad scale trends between different phyla.

Results

The NR reference phylogeny shows many expected relationships between the phytoplankton species, along with some surprising results (Fig. 5). There is a polytomy at the first node in the tree as a result of low bootstrap values. Five NR sequences from unidentified species (Table 3) plotted onto tree in a variety of locations, including, notably, the sequences from three unidentified species sorting into a monophyletic group with each other. All phytoplankton clades on the tree belong to six different phyla (Table 4). All the cryptomonads

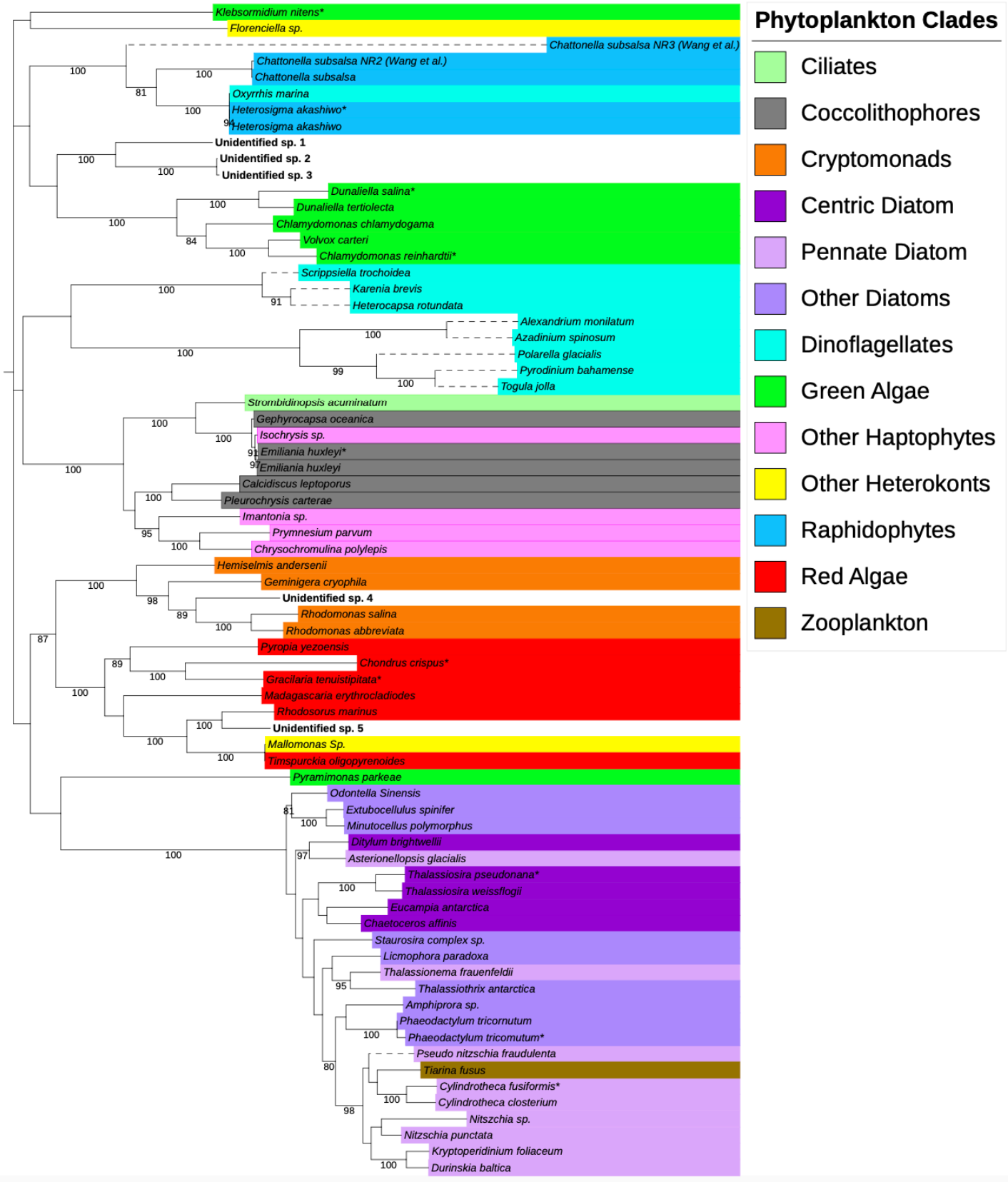

Figure 5. Nitrate Reductase Reference Phylogeny. Constructed using NR amino acid sequences for each species, with Maximum Likelihood method and bootstrap run 100 times for support. Nodes with bootstrap values < 10% were collapsed. Species used to build initial Hmm profile are denoted with asterisks next to their names. Species with NR3 or NR3-like isoforms are denoted with dashed branches.

sorted into a monophyletic group with the addition of unknown species sequence 4. The Raphidophytes sorted into a monophyletic group with the addition of heterotrophic dinoflagellate *Oxyrrhis marina*. The other eight dinoflagellate species (all autotrophic) sorted into a monophyletic group. All diatoms sorted into a monophyletic group with the single zooplankton included, *Tiarina fusus*. The rest of the major phytoplankton clades were paraphyletic. Two heterotrophic plankton (*Oxyrrhis marina* and *Tiarina fusus*) had homologous NR sequences that plotted onto the reference phylogeny, despite the fact that heterotrophs do not have the NR enzyme. The NR3 and NR2 isoforms of *C. subsalsa*, are separated by long branch lengths but the NR3 isoform from *C. subsalsa* did not cluster with other NR3-like sequences, which were found in nine other plankton species (Table 2), including eight dinoflagellates and one diatom.

Table 3. Identification numbers for the strains from the five unidentified species included in the Nitrate Reductase Reference Phylogeny.

Unidentified Species	Identification numbers
Unidentified Sp. 1	Unidentified_sp_Strain_CCMP2111_tax1461542_locID_CAMPEP_0198235342_seqID16412359/77-875
Unidentified Sp. 2	Unidentified_sp_Strain_RCC1871_tax1461544_locID_CAMPEP_0198469254_seqID16520937/86-894
Unidentified Sp. 3	Genus_nov_species_nov_Strain_RCC2335_tax464262_locID_CAMPEP_0197514584_seqID14807595/12-801
Unidentified Sp. 4	non_described_non_described_Strain_CCMP2293_tax697907_locID_CAMPEP_0180150428_seqID11895051/22-860
Unidentified Sp. 5	Unidentified_sp_Strain_CCMP1999_tax100272_locID_CAMPEP_0198726546_seqID16656322/68-888

Table 4. Clades listed in phylogenies organized by what phylum then are contained in, with the phyla used in heat maps. Ochrophyta was excluded from heat maps due to the lack of abundance shown in the eRNA placement fat plot.

Phylum	Clades Contained
Alveolata	Ciliates, Dinoflagellates
Archaeplastida	Green Algae, Red Algae
Cryptomonads	Cryptomonads
Haptophyta	Coccolithophores, Other Haptophytes
Ochrophyta	Centric Diatoms, Pennate Diatoms, Other Diatoms
Stramenopiles	Other Heterokonts, Raphidophytes

The Environmental RNA placement 'fat plot' reveals the rough distribution of which species and clades were present and actively transcribing Nitrate Reductase in the transition zone of the North Pacific Subtropical Gyre during Gradients 1 and 2 (Fig. 6). Dinoflagellates, raphidophytes, coccolithophores, other haptophytes, and some red algae species were abundant and actively transcribing the NR gene. In addition, some species of green algae, specifically *Klebsormidium nitens* and *Pyramimonas parkeae* were also in high abundance and actively transcribing NR. Unidentified species 1-3 were all present and actively transcribing, though RNA transcripts indicate in lower abundance. Cryptomonads were present and sampled on the cruise, though in lower abundance than the clades previously mentioned. Finally, several species of diatoms were detected including *Pseudo nitzchia fraudulenta* and both *Thalassiosira sp.*, but they had much lower abundance than the other clades. The other diatom species had few (<10) or no RNA transcripts plot onto their branches of the phylogeny, which indicates that they were either absent or present but not actively transcribing the gene coding for NR.

The nitrate plus nitrite concentrations sampled on the Gradients 1 cruise (Fig. 7A) were low (<0.01 μM) until 33.3ºN, when concentration increased to approximately 1μM. Both stations

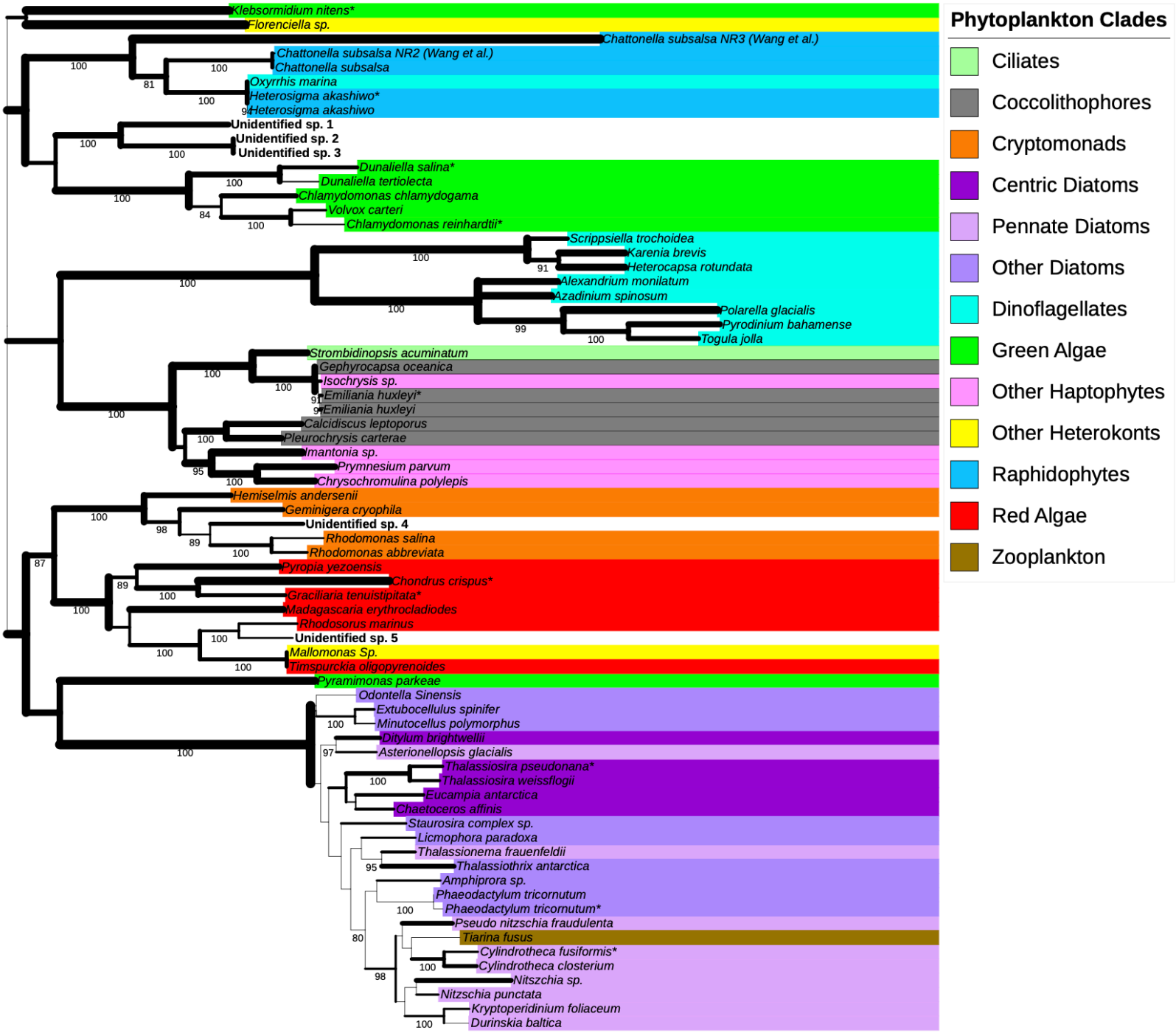

Figure 6. Environmental RNA placement on top of NR Reference Phylogeny. Environmental RNA data is taken from both Gradients 1 and 2 cruises. Bootstrap values >80% are listed below branches. Thickness of the branch is proportional to the number of RNA sequences that plotted on the branch. The thickest branch has 1020 RNA transcripts plotted onto it. The thinnest branches (e.g. those to some diatoms species) do not have any RNA transcripts plotted onto them.

sampled further north, at 36.6ºN and 37.3ºN, have high concentrations of nitrate, approximately 5.7µM and 5.9µM respectively. The heat maps of RNA transcripts from 0.2µm-3µm sized cells sampled on Gradients 1 (Fig. 8, Fig. 9) show that Haptophytes had high abundance at 26.3ºN and 37.3ºN. The majority of Stramenopiles had high abundance at 35.5ºN, and moderate abundance at either 32.6ºN or 33.1ºN. Alveolata groups were more variable, but most had high or moderate abundance at 32.6ºN and 33.1ºN, with a few clades expressing moderate abundance at 36.57ºN and 37.3ºN. Archaeplastida had overall high abundance between 32.6ºN and 35.5ºN, with some groups expressing moderate abundance at 36.57ºN and 37.3ºN. Cryptomonads had no clear trends, with some groups having moderate to high abundance at 26.3ºN and 28.1ºN, some at 33.1ºN and 35.5ºN, and some at 36.57ºN

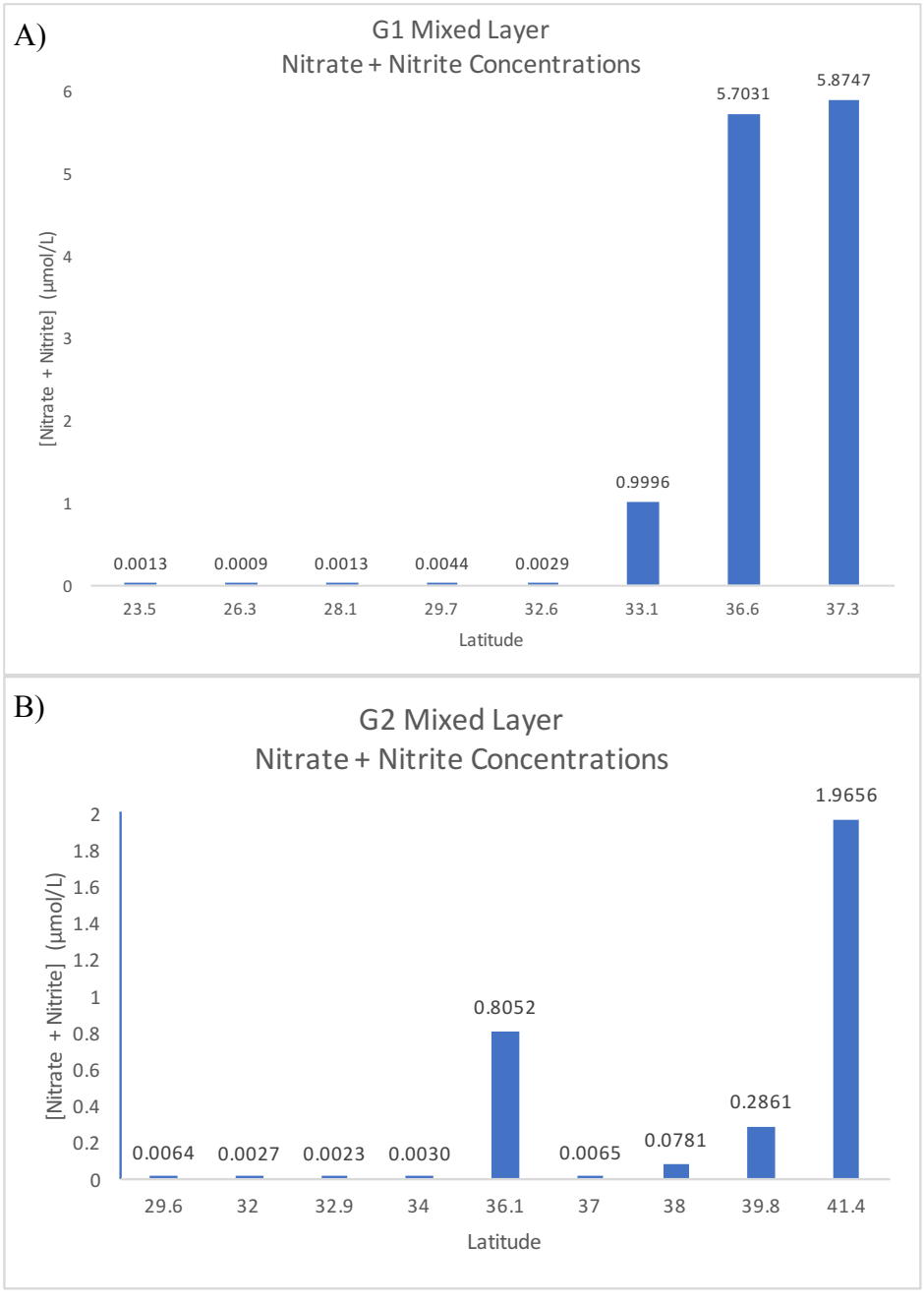

Figure 7. Nitrate plus Nitrite concentrations from (A) Gradients 1 and (B) Gradients 2 cruises, plotted by latitude. Concentrations are in units of µM, and latitudes are degrees north.

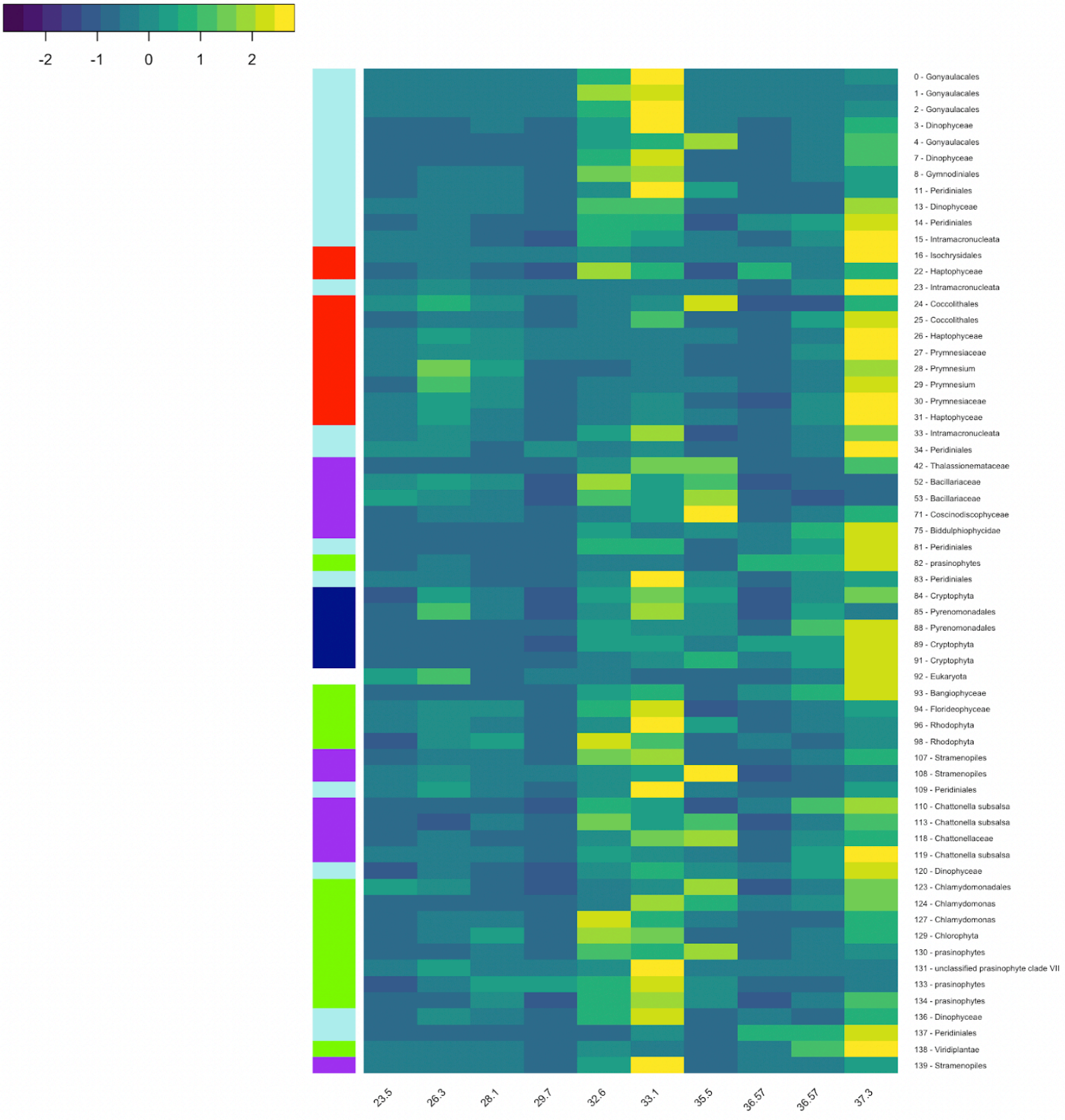

Figure 8. Heat map of the Gradients 1 0.2 µm cell filter RNA transcript abundance. The x-axis corresponds to latitude in degrees north and the y-axis is lowest level clade that could be identified for an RNA transcript. The visualization of the heat map RNA transcript data has been Z-score normalized, so the scale is unitless. The column to the left of the heat map is color coded to the phylum that the clade belonged to. Light blue = Alveolata. Green = Archaeplastida. Dark Blue = Cryptomonads. Red = Haptophyta. Purple = Stramenopiles. The two columns at 36.57ºN were samples taken at the same station, but at two different times.

Figure 9. Heat map of Gradients 1 0.2 µm filter RNA abundance normalized to chlorophyll concentration. The x-axis corresponds to latitude in degrees north and the y-axis is lowest level clade that could be identified to the RNA transcript. The visualization of the heat map RNA transcript data has been Z-score normalized, so the scale is unitless. The column to the left of the heat map is color coded to the phylum that the clade belonged to. Light blue = Alveolata. Green = Archaeplastida. Dark Blue = Cryptomonads. Red = Haptophyta. Purple = Stramenopiles. The two columns at 36.569ºN and 36.57ºN were samples taken at the same station, but at two different times.

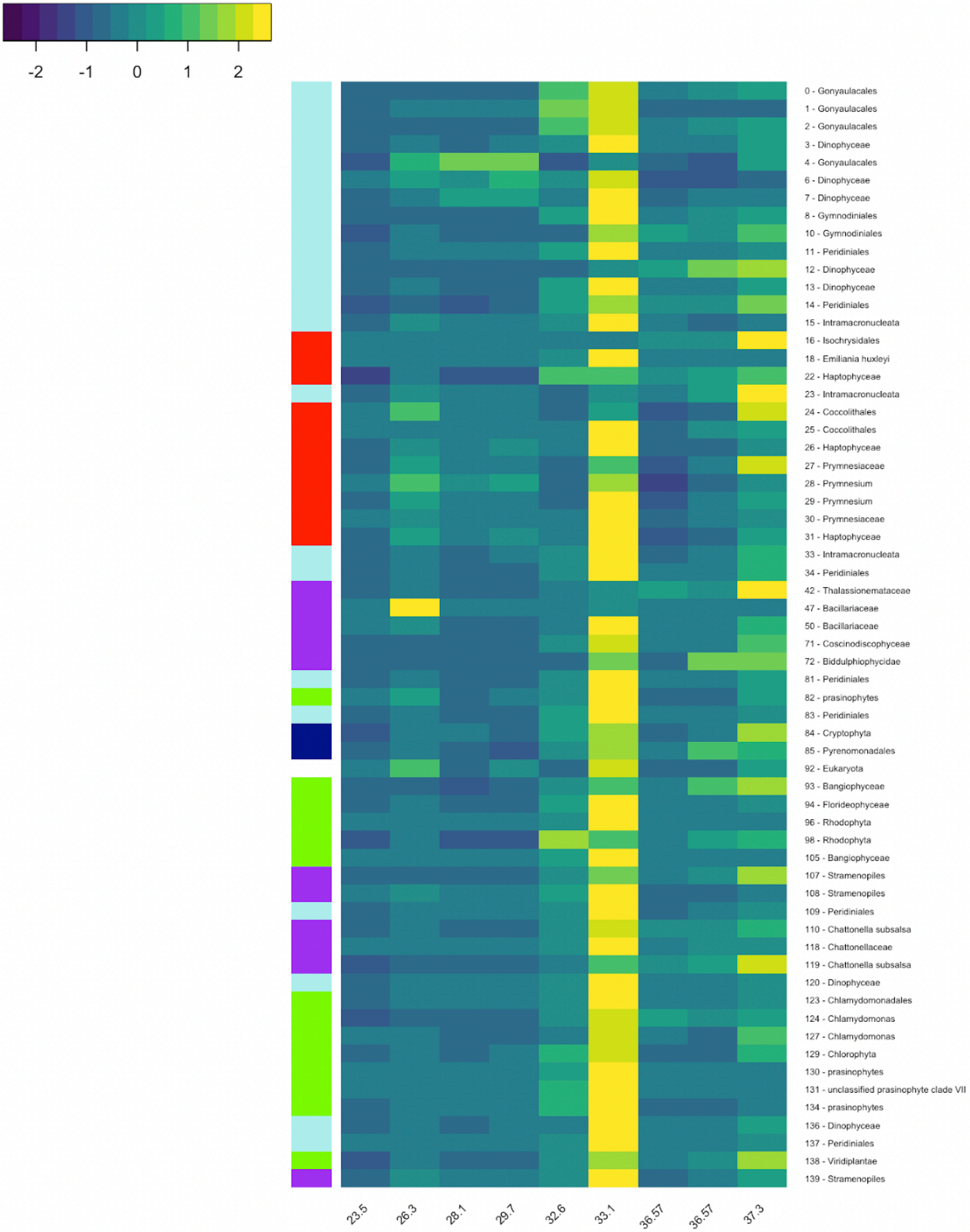

Figure 10. Heat Map of Gradients 1 3µm filter RNA abundance. The x-axis corresponds to latitude in degrees north and the y-axis is lowest level clade that could be identified to the RNA transcript. The visualization of the heat map RNA transcript data has been Z-score normalized, so the scale is unitless. The column to the left of the heat map is color coded to the phylum that the clade belonged to. Light blue = Alveolata. Green = Archaeplastida. Dark Blue = Cryptomonads. Red = Haptophyta. Purple = Stramenopiles. The two columns at 36.57ºN were samples taken at the same station, but at two different times.

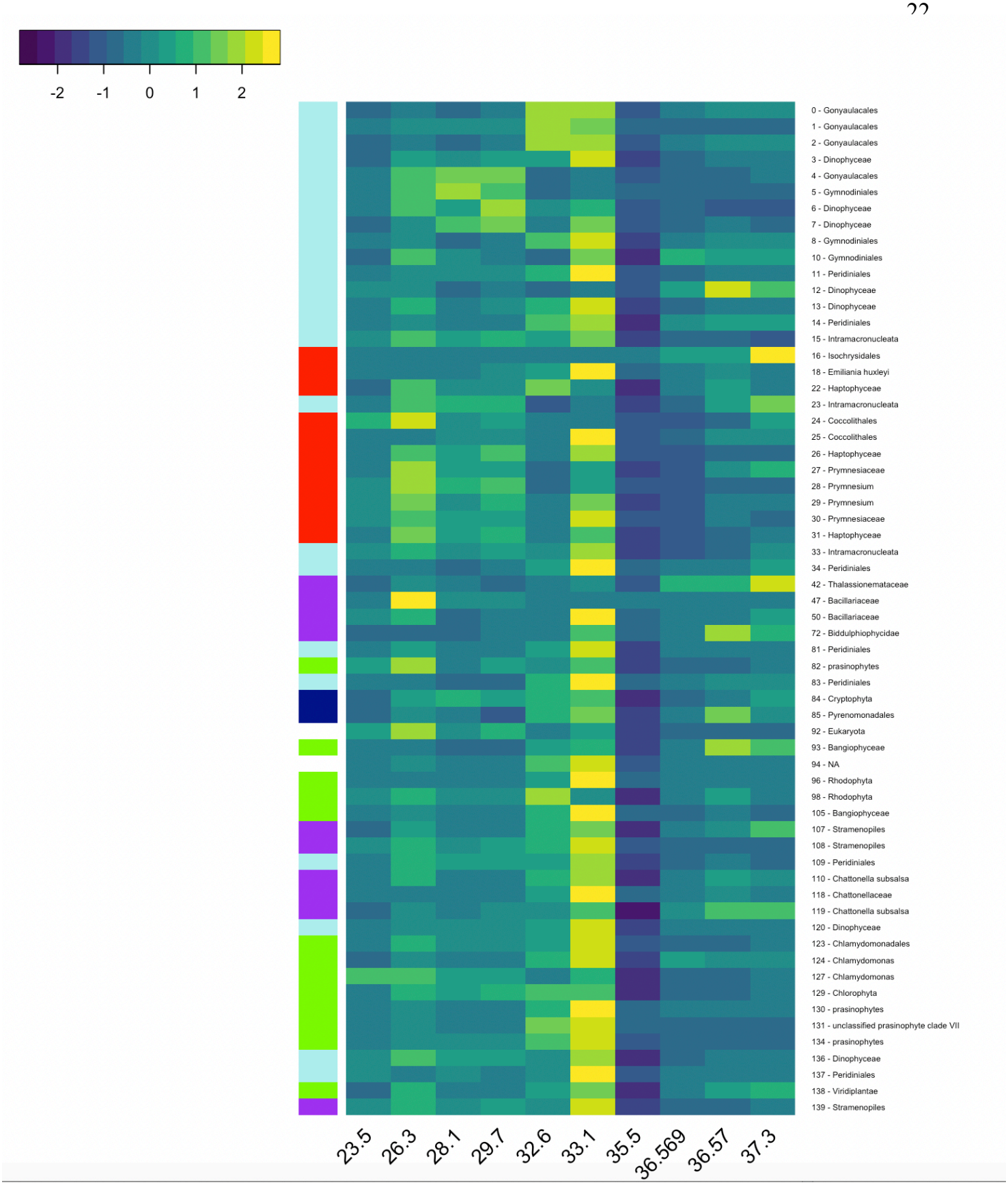

Figure 11. Heat map of Gradients 1 3µm filter RNA abundance normalized to chlorophyll concentration. The x-axis corresponds to latitude in degrees north and the y-axis is lowest level clade that could be identified to the RNA transcript. The visualization of the heat map RNA transcript data has been Z-score normalized, so the scale is unitless. The column to the left of the heat map is color coded to the phylum that the clade belonged to. Light blue = Alveolata. Green = Archaeplastida. Dark Blue = Cryptomonads. Red = Haptophyta. Purple = Stramenopiles. The two columns at 36.569ºN and 36.57ºN were samples taken at the same station, but at two different times.

and 37.3ºN. These trends are present in both the unnormalized and chlorophyll normalized data.

By comparison, the RNA transcripts from the 3-100µm size cells sampled during Gradients 1 show a very different trend, with low abundance of the majority of the clades at the more southern latitudes (latitudes less than 33ºN) (Fig. 10, Fig. 11). The majority of groups from all phyla (Alveolata, Archaeplastida, Cryptomonads, Haptophyta, and Stramenopiles) had their highest abundance at 33.1ºN. There was low abundance from the majority of groups in all phyla at 35.5ºN. A few select clades from Alveolata, Archaeplastida, Haptophyta, and Stramenopiles have moderate or high abundance at 36.57ºN and 37.3ºN. These trends are present in both the unnormalized and chlorophyll normalized data.

The nitrate plus nitrite concentrations sampled during Gradients 2 were less than 0.1µM at all but three stations (Fig. 7B). The concentration at 36.1ºN was approximately 0.8µM, at 39.8ºN it was approximately 0.29µM, and at 41.4ºN it was approximately 1.97µM. The highest concentration of nitrate observed during Gradients 2 was much lower than that observed during Gradients 1 (1.97µM versus 5.9µM) (Fig. 7). In the unnormalized figures, the RNA from 0.2-3µm sized cells is seen in higher abundance at 36.1ºN and then at latitudes greater than 38ºN, with no clear distinction in the trend between phyla (Fig. 12). By comparison, in the chlorophyll normalized data, we see that Cryptomonads is most abundant at the furthest north of any of the phyla, 36.1ºN, and all phyla but Cryptomonads had high abundance at 25.8ºN (Fig. 13). Some groups from Archaeoplastida, Alveolata, and Haptophyta also have moderate or high abundance between 32.9ºN and 36.1ºN.

In the unnormalized RNA data from 3-100µm sized cells groups in all phyla have moderate and high abundance at 39.8ºN and 41.4ºN, and a select few groups from Alveolata and

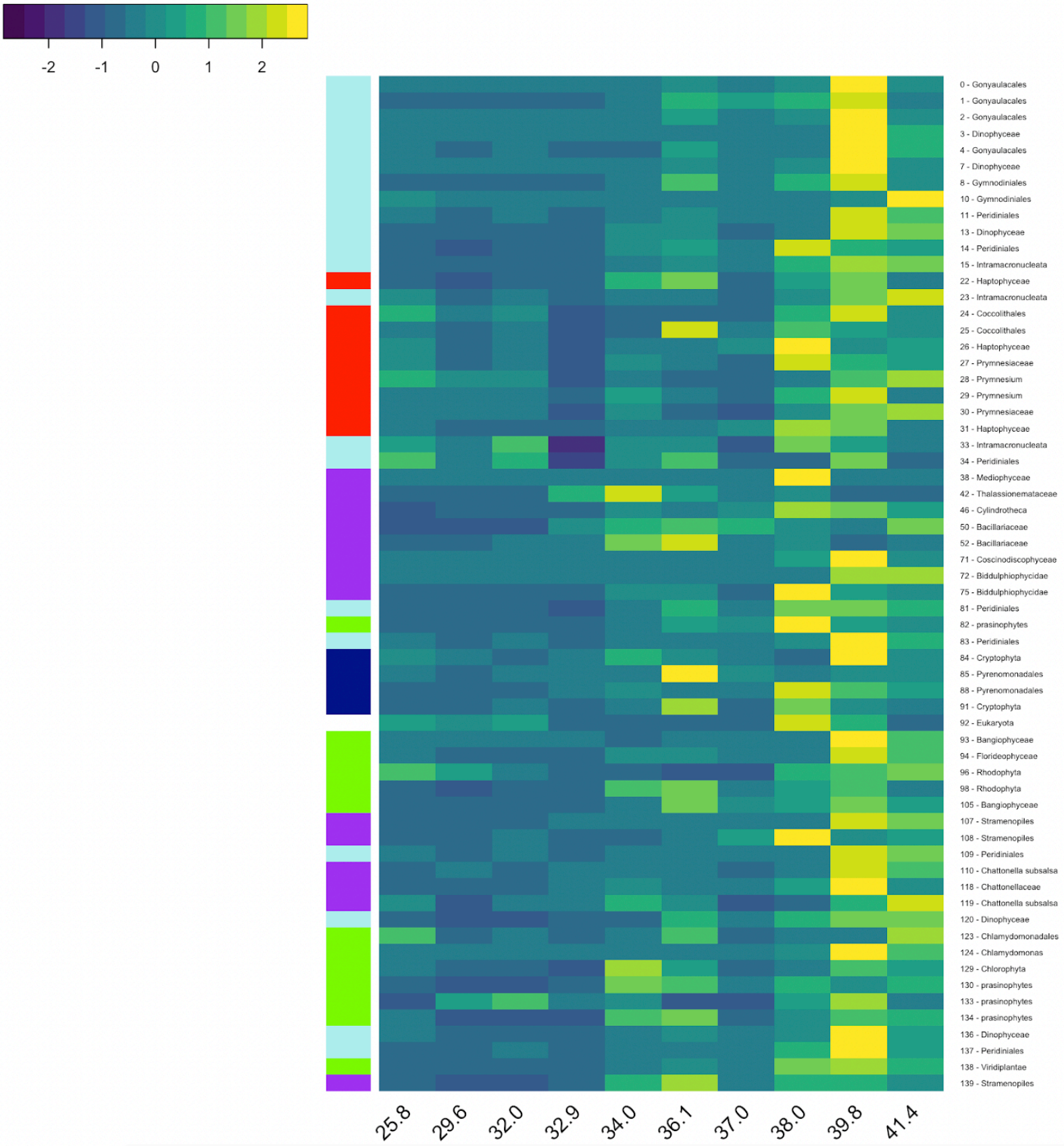

Figure 12. Heat map of Gradients 2 0.2µm filter RNA abundance. The x-axis corresponds to latitude in degrees north and the y-axis is lowest level clade that could be identified to the RNA transcript. The visualization of the heat map RNA transcript data has been Z-score normalized, so the scale is unitless. The column to the left of the heat map is color coded to the phylum that the clade belonged to. Light blue = Alveolata. Green = Archaeplastida. Dark Blue = Cryptomonads. Red = Haptophyta. Purple = Stramenopiles.

Figure 13. Heat map of Gradients 2 0.2µm filter RNA abundance normalized to chlorophyll concentration. The x-axis corresponds to latitude in degrees north and the y-axis is lowest level clade that could be identified to the RNA transcript. The visualization of the heat map RNA transcript data has been Z-score normalized, so the scale is unitless. The column to the left of the heat map is color coded to the phylum that the clade belonged to. Light blue = Alveolata. Green = Archaeplastida. Dark Blue = Cryptomonads. Red = Haptophyta. Purple = Stramenopiles.

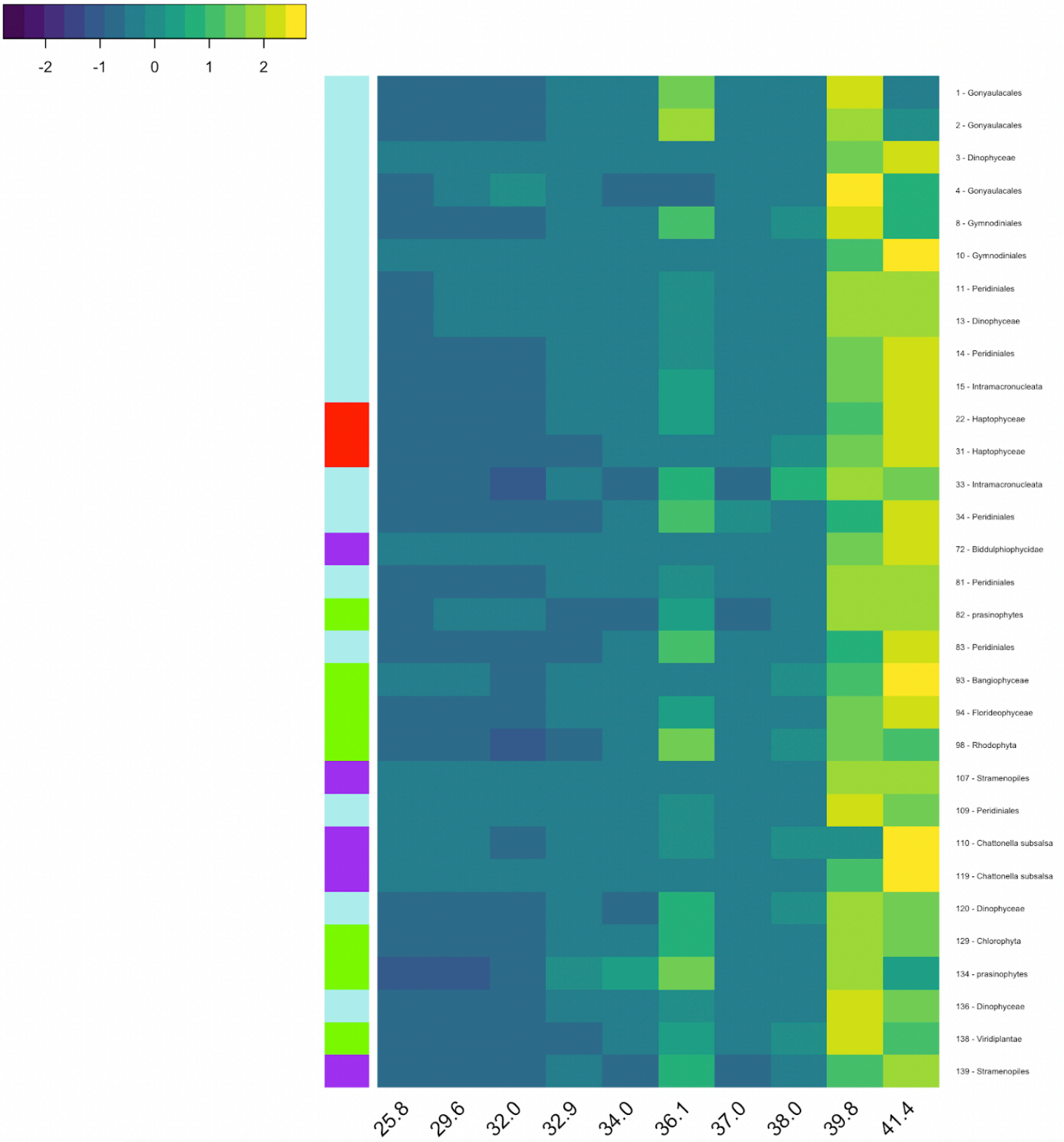

Figure 14. Heat map of Gradients 2 3µm filter RNA abundance. The x-axis corresponds to latitude in degrees north and the y-axis is lowest level clade that could be identified to the RNA transcript. The visualization of the heat map RNA transcript data has been Z-score normalized, so the scale is unitless. The column to the left of the heat map is color coded to the phylum that the clade belonged to. Light blue = Alveolata. Green = Archaeplastida. Dark Blue = Cryptomonads. Red = Haptophyta. Purple = Stramenopiles.

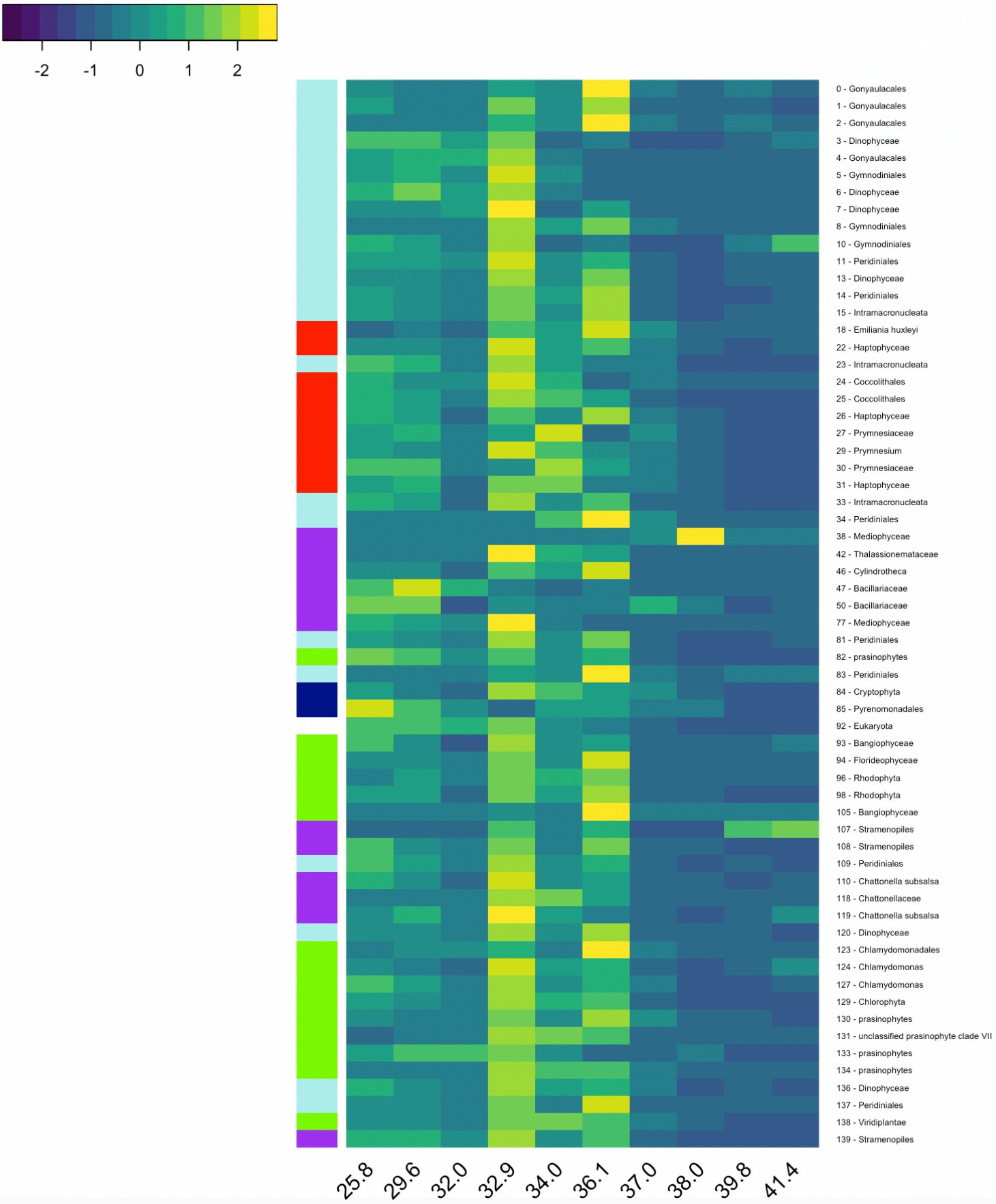

Figure 15. Heat map of Gradients 2 3µm filter RNA abundance normalized to chlorophyll concentration. The x-axis corresponds to latitude in degrees north and the y-axis is lowest level clade that could be identified to the RNA transcript. The visualization of the heat map RNA transcript data has been Z-score normalized, so the scale is unitless. The column to the left of the heat map is color coded to the phylum that the clade belonged to. Light blue = Alveolata. Green = Archaeplastida. Dark Blue = Cryptomonads. Red = Haptophyta. Purple = Stramenopiles.

Archaeplastida have moderate abundance at 36.1ºN (Fig. 14). By comparison, the chlorophyll normalized data show that only 3 different clades have moderate or high abundance at 38ºN or any latitude further north (Fig. 15). Instead, all phyla have moderate to high abundance between 32.9ºN and 36.1ºN. In addition, a handful of clades from Cryptomonads, Stramenopiles, and Haptophyta had moderate abundance at 25.8ºN and 29.6ºN.

Discussion

Nitrate Reductase Reference Phylogeny

The RNA transcripts for the gene encoding Nitrate Reductase were used to determine the distributions of different phytoplankton species. A number of unexpected findings were discovered, including the Nitrate Reductase sequences from two heterotrophic plankton plotting onto the NR phylogeny (Fig. 5). The simplest explanation for this result is that the two species *Oxyrrhis marina* and *Tiarina fusus* consumed phytoplankton shortly before being sampled and sequenced. These NR sequences used in phylogeny construction likely came from the phytoplankton they consumed. This also explains where these sequences group on the phylogeny. *Oxyrrhis marina* is a dinoflagellate, and yet the NR clusters within Raphidophyte plankton, which suggests that *O. marina* likely consumed a Raphidophyte, and based off of the polytomy with both *H. akashiwo* sequences, most likely had consumed a *H. akashiwo* shortly before being sequenced. Similarly, the *Tiarina fusus* NR clusters with pennate diatoms, and sorts into a monophyletic group with both *Cylindrotheca fusiformis* sequences, suggesting *T. fusus* had likely consumed a diatom, most likely a *C. fusiformis,* shortly before being sampled and sequenced.

The fact that the eight dinoflagellate species and *Pseudo nitschia fraudulenta* that had gaps in their NR sequence similar to the *C. subsalsa* NR3 isoform did not plot on the NR reference phylogeny near the NR3 isoform (Fig. 5). This indicates that either they did not encode a true NR3 isoform, or that they do encode the NR3 isoform but there are other parts of the NR sequence that are more important in determining their phylogenetic relationships. However, this result suggests that there could be other species besides *C. subsalsa* that have the NR3 isoform. More research will have to be done to determine if this is the case.

Five NR sequences from unidentified species plotted onto the NR phylogeny (Fig. 5). Three NR sequences (1-3) clustered in a monophyletic group with each other that then clustered with five green algae species, suggesting that these sequences are derived from green algae species that have not had their NR sequence characterized yet. Unidentified species 4 clustered within the middle of the monophyletic Cryptomonads, which means it is likely that the sequence comes from a new species of Cryptomonad, or a closely related species that has not had its NR sequence characterized yet. Unidentified species 5 NR sequence clustered within the monophyletic group of red algae plus another heterokont (*Mallomonas sp.*), which supports that it is likely a red algae. Future research will have to study these sequences more closely to see if there is any merit to these hypotheses and if species identification can be done.

Environmental RNA Fat Plot

The lack of RNA transcripts placed on branches leading to diatom species seems to indicate that there were either no diatom species sampled during Gradients 1 and 2 or that the diatom species were present but not actively transcribing NR (Fig. 6). However, the single

branch on the tree leading to the monophyletic group of all diatoms had the most RNA transcripts plotted on the fat plot (1020 transcripts). This likely means that the reference data base lacks diatom species related to those found in the environmental RNA samples. If this is the case, then there were actually diatom species present and transcribing NR, but there is no way to differentiate between diatom species and get further resolution on which species and diatom types (e.g pennate vs. centric) in particular were abundant.

Another interesting feature highlighted by the environmental RNA placement is that RNA transcripts were associated with *Polarella glacialis*, a cold-adapted dinoflagellate that is typically only found in the Southern ocean and around the Arctic. This indicates a species with a NR sequence closely related to the NR sequence of *P. glacialis* was present in the environmental data. More research will need to be done in the future to indicate whether this was an anomalous sampling, or if a dinoflagellate similar to *P. glacialis* was found in the temperate waters north of Hawaii, and has a more cosmopolitan distribution than previously thought for cold-adapted phytoplankton.

Heat Maps

It has been well established in oceanography research that large phytoplankton are found in higher nitrate concentrations while smaller phytoplankton are found in oligotrophic environments, as size will change the surface area to volume ratio for a cell, which determines their maximum rate of nutrient uptake (Stolte and Riegman, 1995). The larger a cell is, the lower the ratio, which means there is less surface area relative to the volume of the organism, and therefore the higher the environmental nutrient concentration the cell needs to meet their

metabolic requirements. Despite this, the unnormalized heat maps (Fig. 8, 10, 12, 14) show that all phytoplankton, no matter what size or phylum, had higher abundance at the northernmost latitudes on both cruises where nitrate concentrations were significantly higher than at southern latitudes. This indicates that all phytoplankton, even those which can survive in oligotrophic environments, will thrive in higher nutrient systems.

Table 5. Table of latitudes from each cruise and cell size at which the majority of clades from a phylum were most abundant, along with the nitrate concentration measured at that latitude. CD = Clade Dependent (No broad trends across majority of clades in phylum). NA = Not Applicable. NA put in for nitrate concentrations when nitrate was not sampled at that latitude.

	0.2µm filter				3µm filter			
Phylum	**G1 CN**	**Nitrate**	**G2 CN**	**Nitrate**	**G1 CN**	**Nitrate**	**G2 CN**	**Nitrate**
Alveolata	32.6ºN 33.1ºN	0.0029µM 0.9996µM	25.8ºN 36.1ºN	NA 0.8052µM	33.1ºN	0.9996µM	32.9ºN 36.1ºN	0.0023µM 0.8052µM
Archaeplastida	32.6ºN 33.1ºN 35.5ºN	0.0029µM 0.9996µM NA	25.8ºN 32.9ºN 34ºN	NA 0.0023µM 0.003µM	33.1ºN	0.9996µM	32.9ºN 36.1ºN	0.0023µM 0.8052µM
Cryptomonads	CD	NA	CD	NA	CD	NA	CD	NA
Haptophyta	26.3ºN 37.3ºN	0.0009µM 5.8747µM	25.8ºN	NA	26.3ºN 33.1ºN	0.0009µM 0.9996µM	32.9ºN 34ºN 36.1ºN	0.0023µM 0.003µM 0.8052µM
Stramenopiles	32.6ºN 35.5ºN	0.0029µM NA	32.9ºN 34ºN	0.0023µM 0.003µM	33.1ºN	0.9996µM	32.9ºN	0.0023µM

Using this knowledge of cell size affecting nutrient uptake, it was also expected that large cells (those caught in the 3µm filter) would be found at different latitudes, those further north with higher nitrate concentrations, than the small cells (those caught in the 0.2µm filter). However, comparing what latitudes each phyla was found when between the large and small cell sizes when chlorophyll normalized (Table 5) indicates that both large and small cells in the phyla

Alveolata, Archaeplastida, the Gradients 1 data for Haptophyta, and the Gradients 2 data for Stramenopiles were found at high abundance in similar latitudes and nitrate concentrations.

A possible explanation for this unusual result could be found by looking at flux of nitrate through a system rather than total nitrate concentrations. If the flux of nitrate through a system at any given time is high, then the total nitrate concentration in seawater at any given time could be low because just as nitrate gets added to a system (e.g through upwelling), it gets taken up by phytoplankton and removed from the seawater. This could explain why large phytoplankton were abundant at latitudes with low total nitrate concentration. As long as there was high flux of nitrate, there would still be sufficient nitrate in the low total nitrate systems to support the metabolic needs of these large cells. Future research should be conducted to not only see if large phytoplankton are abundant in other low nitrate concentration regions, but to also measure nitrate flux and see if a high nitrate flux is supporting these cells.

Another possible explanation is that these results are biased as a result of normalizing with respect to total chlorophyll concentration sampled. This normalization method allows for normalizing with respect to estimated total phytoplankton biomass, and visualizing how large of a proportion of the phytoplankton population at a station was comprised of a single phytoplankton clade. Yet this can lead to inaccurate results due to the fact that chlorophyll is not an extremely accurate indicator of phytoplankton biomass since chlorophyll abundance in a cell is not constant between species, and therefore if there was a species of plankton with high chlorophyll in their cells, biomass would be overestimated and possible make normalized RNA transcript abundance look lower than it is. Future studies should account for this by normalizing using another method, such as by total RNA transcript abundance per clade.

In addition to cells of different sizes being found at similar latitudes, the phyla Alveolata, Archaeplastida, and Stramenopiles were all found at strikingly similar latitudes and nitrate concentrations to each other across both cell sizes and cruises. Alveolata is a phylum of plankton (both phototrophic and heterotrophic) that includes both ciliates and dinoflagellates (Leander and Keeling, 2004). Archaeplastida is a phylum of eukaryotes which contain the photoautotrophic red algae and green algae (Collén et al., 2013). Stramenopiles is a diverse phylum consisting of both unicellular and multicellular eukaryotic algae which are denoted by the presence of tripartite hairs in their flagella (Massana et al., 2004). The fact that these strikingly different phyla were all found in similar nitrate concentrations indicates that they have similar distributions at the phylum level in the North Pacific. More research should be done to see if this trend is maintained in other ocean basins. In addition, future research should look at the distributions of organisms within these phyla at a lower taxonomic level (e.g family or genus) to see if this pattern is maintained or changes when looking at more specific organisms.

Cryptomonad is a phylum of unicellular plankton which can be found anywhere from fresh to marine water systems (Hoef-Emden et al., 2002). Very few cryptophytes (also called cryptomonads) were identified in this study, with a maximum of five clades of cryptophytes seen in any given RNA transcript data. However, the few cryptophytes that were present displayed different abundances at different latitudes and nitrates concentrations which prevented any trends in distribution being determined at the phylum level (Table 4). For small cells in Gradients 1, one clade was found in high abundance only at 26.3ºN, another one was found in high abundance only at 35.5ºN, and the two were found only in moderate abundance at 36.57ºN and 37.3ºN. In Gradients 2 one clade was found in high abundance at 25.8ºN, another at 34ºN, and two more

had high abundance at 36.1ºN. This indicates that some small cryptophytes favor oligotrophic environments with low nitrate, others favor moderate nitrate, while other still favor high nitrate environments. More research will have to be done with a focus on cryptophytes to see what family and genera favor different nitrate abundances.

The large cryptophytes present an even more interesting case. After chlorophyll normalization, large cryptophytes, no matter the latitude, were never a large proportion of the total phytoplankton biomass in the North Pacific. The two large cryptophyte clades that plotted on the Gradients 1 and 2 heat maps did have low to moderate abundance at a variety of latitudes, but were never in high abundance. In Gradients 1 one clade did show moderate abundance at 36.57ºN, and in Gradients 2 a different clade showed moderate abundance at 25.8ºN, indicating that large cryptophytes, similarly to their small counterparts, have a diversity in what nitrate concentrations they can survive in. Future research will have to look deeper into large cryptophytes in the North Pacific to not only look at the family and genus level to see which clades favor different nitrate concentrations, but also to determine why they never reach high abundance, such as top-down and bottom-up controls or being outcompeted by other phytoplankton species.

Haptophyta is a phylum consisting of eukaryotic phytoplankton, a well known example being coccolithophores (Burki et al., 2012). The haptophytes found in this study deviated significantly from the distribution trends seen in Alveolata, Archaeplastida, and Stramenopiles. The chlorophyll normalized data from Gradients 1 shows small haptophytes made up a significantly large proportion of phytoplankton found in extremely low and extremely high nitrate concentrations, and while the small haptophytes in Gradients 2 were sampled at a latitude

where nitrate was not also sampled (25.8ºN), it can be assumed that the relatively southern latitude of 25.8ºN likely had low nitrate concentrations similar to stations sampled at nearby latitudes (Table 4). The large haptophytes sampled on Gradients 1 and 2 were found at the same latitudes as other large phytoplankton, but were also found in significantly high abundance at very low nitrate concentrations (Table 4). While finding these large haptophytes in a low total nitrate environment could be the result of a large nitrate flux occurring at those latitudes as discussed earlier, putting it together with the findings for small haptophytes indicates that this phylum as a whole consists of species that have evolved to thrive in oligotrophic and low nitrate environments. However, there is a low likelihood of this being the case. All latitudes sampled in the cruises were contained within the North Pacific Subtropical Gyre, which is a downwelling system. Because it is a downwelling system, a low nitrate flux is much more likely since there is no direct influx of high nitrate waters as is seen in upwelling systems. The more probable explanation is that the large haptophytes are mixotrophic, such that they can rely on consuming other organisms when nutrient levels are too low to support photoautotrophy. The haptophytes plus a single ciliate plotted into a monophyletic group on the nitrate reductase reference phylogeny (Fig. 3), so more research should be done on their nitrate reductase sequences to determine what common feature of their NR enzymes allows them to thrive in oligotrophic environments and outcompete other species, or to see if it is not a unique feature of NR but rather an ability to be mixotrophic.

Conclusions

This study was set out to determine if there was a strong correlation between the abundance of RNA transcripts for Nitrate Reductase and the total concentration of nitrate in the environment to evaluate whether ambient nitrate be used to predict the location of a phytoplankton taxa. In addition, this study evaluated the similarity of Nitrate Reductase amino acid sequences between a variety of phytoplankton species to see if clades naturally sort into monophyletic groups, allowing for conclusions about their distribution and NR characteristics to be made across the clades. Nitrate concentrations are known to control distribution of phytoplankton species as different species utilize different ecological niches with varying nitrate availability. Nitrate and overall nutrient requirements for phytoplankton are also determined by the size of phytoplankton, with larger phytoplankton requiring higher nitrate conditions. In addition, NR enzyme follows regular Michaelis-Menten kinetics, which means that in sufficiently high nitrate concentrations, the only thing limiting the maximum possible nitrate uptake by a phytoplankton is the maximum activity of NR, which is dictated by its primary amino acid sequence. With this knowledge, it was hypothesized that major clades of phytoplankton would sort into monophyletic groups based on their NR primary sequence, and that these clades would have distinct distributions of abundance determined by environmental nitrate concentrations, such that phytoplankton distributions could be predicted based on the ambient nitrate concentration.

Undescribed NR3-like isoforms were present in nine phytoplankton species. Future studies should look into whether the NR3-like isoforms found in the eight dinoflagellates and one diatom species identified here, along with NR sequences from other clades, function as true

NR3 isoform, as this could fundamentally change our understanding of Nitrate Reductase and its conserved domains required for function. In addition, five unidentified NR sequences plotted onto the phylogeny. More in-depth phylogenetic analysis should be conducted in the future to determine what taxonomic groups these sequences belong to.

The RNA transcript heat maps show there are distinct trends in distribution at the phylum level, with several phyla (Alveolata, Archaeplastida, and Stramenopiles) having similar trends in distribution while Haptophytes were distinct in making up a large proportion of phytoplankton found in extremely oligotrophic environments. However, the lack of a clear pattern in the phylum Cryptomonad shows that some phyla have enough diversity within them that the distribution patterns should be determined at a lower taxonomic level, such as family or genus, to see whether clear trends emerge. In addition, large and small phytoplankton were found in similar latitudes and nitrate concentrations when RNA transcript data was normalized by chlorophyll, which could indicate that along with total nitrate concentration, nitrate flux might play an important role in determine distribution of cells based on size.

More studies like this will need to be done in other ocean basins to see if similar results are obtained, or if these patterns are unique to the North Pacific. In addition, more diverse NR sequences need to be added to reference databases so future research can obtain a higher taxonomic resolution to elucidate patterns of distribution at the family and genus level. Future studies should also quantify the flux of nitrate through a system to see if large nitrate flux correlates with abundance of large phytoplankton as a possible explanation for why large phytoplankton are detected in low nitrate environments.

Climate change has been and will continue to alter oceanic conditions such as with ocean warming and increased ocean stratification leading to decreased nutrient availability for phytoplankton (Steinacher et al., 2010). Phytoplankton not only play a critical role in marine ecosystems, but act as the base food source for fisheries worldwide (Howell et al., 2013), which means the negative impacts of climate change will be felt by people and other organisms across globe. This alone provides a critical reason for focusing future research on quantifying the oceanic factors controlling phytoplankton distribution, as fisheries play key roles worldwide in global economies providing food sources for people along with jobs. Continuing research can discover the key factors controlling phytoplankton distribution in the ocean. This knowledge can be used to build models of how phytoplankton communities will be altered as climate change continues, and inform marine conservation policies to help protect phytoplankton and all the organisms that rely on them.

Acknowledgements

I would first like to thank Dr. E. Virginia Armbrust for mentoring me through the conception, designing, and execution of this project. I would also like to offer my sincerest thanks to Dr. Sacha Coesel for her significant help to me in writing code, conducting my statistical analyses, and answering all my emails full of questions as I learned how to write code in R. I would also like to thank my parents and older brother for supporting my scientific endeavors.

References

Allen, A. E., Ward, B. B., & B. Song. 2005. Characterization of diatom (*Bacillariophyceae*) nitrate reductase genes and their detection in marine phytoplankton communities. J. Phycol. 41(1): 95–104.

Baldauf, S. L. 2003. The deep roots of eukaryotes. Sci. 300(5626): 1703–1706.

Barton, A. D., Irwin, A. J., Finkel, Z. V., & C. A. Stock. 2016. Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. Proc. Natl. Acad. Sci. U.S.A.. 113(11): 2964–2969.

Burki, F., Okamoto, N., Pombert, J. F., & P. J. Keeling. 2012. The evolutionary history of haptophytes and cryptophytes: Phylogenomic evidence for separate origins. Proc. Royal Soc. B. 279(1736): 2246–2254.

Brown, C. J., Fulton, E. A., Hobday, A. J., Matear, R. J., Possingham, H. P., Bulman, C., Christensen, V., Forrest, R. E., Gehrke, P. C., Gribble, N. A., Griffiths, S. P., Lozano-Montes, H., Martin, J. M., Metcalf, S., Okey, T. A., Watson, R., & A. J. Richardson. 2010. Effects of climate-driven primary production change on marine food webs: Implications for fisheries and conservation. GCB. 16(4): 1194–1212.

Campbell, W. H.. 1999. Nitrate reductase structure, function and regulation: Bridging the gap between biochemistry and physiology. Annu. Rev. Plant. Biol. 50:277–303.

Chavez, F. P., Barber, R. T., Kosro, P. M., Huyer, A., Ramp, S. R., Stanton, T. P., & B. Rojas De Mendiola. 1991. Horizontal transport and the distribution of nutrients in the coastal transition zone off northern California: effects on primary production, phytoplankton biomass and species composition. J. Geophys. Res. 96(C8): 14833-14848.

Cheung, W. W. L., Lam, V. W. Y., Sarmiento, J. L., Kearney, K., Watson, R., Zeller, D., & D. Pauly. 2010. Large-scale redistribution of maximum fisheries catch potential in the global ocean under climate change. GCB. 16(1): 24–35.

Collén, J., Porcel, B., Carré, W., Ball, S. G., Chaparro, C., Tonon, T., Barbeyron, T., Michel, G., Noel, B., Valentin, K., Elias, M., François Artiguenave, Arun, A., Aury, J. M., Barbosa-Neto, J. F., Bothwell, J. H., Bouget, F. Y., Brillet, L., Cabello-Hurtado, F., … & C. Boyen. 2013. Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. PNAS. 110(13):5247–5252.

Degerlund, M., & H. C. Eilertsen. 2010. Main Species Characteristics of Phytoplankton Spring Blooms in NE Atlantic and Arctic Waters (68º-80ºN). Estuaries Coast. 33(2): 242–269.

Epply, R. W., & W. H. Thomas. 1969. Comparison of Half-Saturation Constants for Growth and Nitrate Uptake of Marine Phytoplankton. J. Phycol. 5: 375–379.

Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O., & F. J. R. Taylor. 2004. The evolution of modern eukaryotic phytoplankton. Sci. 305(5682): 354–360.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., & P. Falkowski. 1998. Primary production of the biosphere: Integrating terrestrial and oceanic components. Sci. 281(5374): 237–240.

Garnier, S. 2018. viridis: Default Color Maps from 'matplotlib'. R package version 0.5.1.

Harrison, P. J., Boyd, P. W., Varela, D. E., Takeda, S., Shiomoto, A., & T. Odate. 1999. Comparison of factors controlling phytoplankton productivity in the NE and NW subarctic Pacific gyres. Prog. Oceanogr. 43(2–4): 205–234.

Hoef-Emden, K., Marin, B., & M. Melkonian. 2002. Nuclear and nucleomorph SSU rDNA phylogeny in the cryptophyta and the evolution of cryptophyte diversity. J. Mol. Evol. 55(2):161–179.

Howarth, R. W., & R. Marino. 2006. Nitrogen as the Limiting Nutrient for Eutrophication in Coastal Marine Ecosystems : Evolving Views over Three Decades. Limnol. Oceanogr. 51(1): 364–376.

Howell, E. A., Wabnitz, C. C. C., Dunne, J. P., & J. J. Polovina. 2013. Climate-induced primary productivity change and fishing impacts on the Central North Pacific ecosystem and Hawaii-based pelagic longline fishery. Clim. Change. 119(1): 79–93.

Leander B. S., & P. J. Keeling. 2004. Early evolutionary history of dinoflagellates and apicomplexans (Alveolata) as inferred from hsp90 and actin phylogenies. J. Phycol. 40(2): 341–350.

Litchman, E., & C. A. Klausmeier. 2008. Trait-based community ecology of phytoplankton. Annu. Rev. Ecol. Evol. Syst. 39: 615–639.

Lukhtanov, V. A. 2010. From Haeckel's phylogenetics and Hennig's cladistics to the method of maximum likelihood: Advantages and limitations of modern and traditional approaches to phylogeny reconstruction. Entomol. Rev. 90(3): 299–310.

Martin, J. H., & R. Michael Gordon. 1988. Northeast Pacific iron distributions in relation to phytoplankton productivity. Deep-Sea Res. 35(2): 177–196.

Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K., & C. Pedrós-Alió. 2004. Phylogenetic and ecological analysis of novel marine stramenopiles. Appl. Environ. Microbiol. 70(6): 3528–3534.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

RStudio Team. 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA

Simmons, M. P., Sudek, S., Monier, A., Limardo, A. J., Jimenez, V., Perle, C. R., Elrod, V. A., Pennington, J. T., & A. Z. Worden. 2016. Abundance and biogeography of picoprasinophyte ecotypes and other phytoplankton in the eastern North Pacific Ocean. Appl. Environ. Microbiol. 82(6): 1693–1705.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9): 1312–1313.

Steinacher, M., Joos, F., Frölicher, T. L., Bopp, L., Cadule, P., Cocco, V., Doney, S. C., Gehlen, M., Lindsay, K., Moore, J. K., Schneider, B., & J. Segschneider. 2010. Projected 21st century decrease in marine productivity: A multi-model analysis. Biogeosciences. 7(3): 979–1005.

Stolte, W., & R. Riegman. 1995. Effect of phytoplankton cell size on transient-state nitrate and ammonium uptake kinetics. Microbiology. 141(5):1221–1229.

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., & B. Venables. 2020. gplots: Various R Programming Tools for Plotting Data. R package version 3.1.1.

Wang, Y., Bouchard, J. N., & K. J. Coyne. 2018. Expression of novel nitrate reductase genes in the harmful alga, Chattonella subsalsa. Sci Rep. 8(1): 1–12.

Wickham, H. 2007. Reshaping Data with the reshape Package. Journal of Statistical Software. 21(12):1-20.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, H., François, R., Henry, L., & Müller, K. 2020. dplyr: A Grammar of Data Manipulation. R package version 1.0.2.