UNIVERSITY OF CALIFORNIA SAN DIEGO

Temporal Dynamics of Marine Microbial Communities at the SIO Pier

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Marine Biology

by

Maitreyi Dnyanadeep Nagarkar

Committee in charge:

Professor Brian Palenik, Chair
Professor Ryan Hechinger
Professor Michael Landry
Professor George Sugihara
Professor Karsten Zengler

2019

The Dissertation of Maitreyi Dnyanadeep Nagarkar is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2019

DEDICATION

To everyone that kept me going.

EPIGRAPH

"When it comes down to it, all of science is …possible." -E. R.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

Many thanks to my advisor, Dr. Brian Palenik, for the time, effort, and resources he has put into getting to know me both as a scientist and as a person. I have appreciated his guidance throughout my six years here and have learned so much from experiencing how he thinks about scientific questions, particularly his penchant for pushing back on all assumptions. I will remember fondly many fun conversations and delicious meals that were had with Brian and Bianca Brahamsha, along with the rest of the lab.

I am very grateful to Emy Daniels for welcoming me to the lab and providing much of my early training – as well as for her friendship during my early years at SIO. I have also been lucky enough to enjoy the company of lab-mates Billy Lambert, Traci Yuen, Karl Hong, Maggie Wang, Entesar Alrubaiaan, Ivan Moreno, Michelle Prieto, and Dr. Sushmitha Vijaya Kumar (among many others). Thanks especially to Entesar for always being a listening ear, and to all my undergraduate mentees for their contributions to many projects. I have also appreciated the support and advice of collaborators Dr. Pete Countway and Dr. Nicole Poulton; Pete in particular has been such an incredible resource and I appreciate his responsiveness and enthusiasm.

Thank you to the members of my committee, each of whom provided unique and helpful feedback that helped shape my thesis. Thank you to the Academic Connections program and the Teaching and Learning commons for allowing my graduate experience to include incredible teaching opportunities, both of which I loved so much. Thanks to Dr. Kelly Goodwin for useful advice and collaboration on a large sampling project. Thanks to Angel Ruacho for help with San Diego bay data and hilarious times in the clean van. Thanks to so many helpful folks in Hubbs

Hall, including (but not limited to) Orna, Dejan, James (x2), Javier, Lavelle, Yvonne, Jackie, and Greg. FM has saved our necks countless times!

Believe it or not, graduate school is hard, but the darkest times were always brightened by many wonderful friends. I can't imagine where I'd be without the "al-gals" – Jesse Traller, Raffie Abbriano, and Sarah Lerch, the latter two of which were amazing officemates. Both made our office a lovely respite from all the challenges of grad school, full of laughter, thoughtful discussions and emotional support. Ali Freibott and Bellineth Valencia have been an endless source of comfort, ideas, and advice to me all throughout my time here. I am so thankful that I always came home to a warm environment, fostered by K.C. and Ty, and later Andy and Karthik. Other members of my cohort have made my time here so much fun, with many hikes, potlucks, and Harry Potter-themed events to keep my morale high. Additionally I have been so fortunate to have an incredibly loving and supportive community outside of San Diego, and many friends with whom I have stayed connected across vast distances and time zones. Allison Tse, Laura Horan, Charles Helms, and Anita Gokhlay (among others) have always been just a phone call away and I appreciate that so, so much.

I don't know if I can adequately express my thanks to my sister Gargie, who has endured a lot of complaining and lengthy monologues (but has, in the meanwhile, perfected the art of the guilt trip). Our antics through the years are some of my most cherished memories, and our shared obsessions were always the most welcome distractions. To my parents, who have raised me with stern generosity and love, I am eternally grateful; I have always felt that I could turn to you for support and encouragement. I looked forward to text messages and delicious care packages from my mother, who is always thinking of others over herself. I turned to my father for advice and conversation on a variety of topics, knowing I could get a new perspective. To my family - I

know that all three of you are always there for me and that means more than anything. Finally, I'm so thankful to Andy Gross, who has been patient, generous, and supportive throughout this most intense experience. Andy, you have helped me stay grounded and lovingly encouraged me to be better in so many ways. But most importantly, you have been permissive of my pizza consumption needs to the greatest extent possible.

There are so many others who played a role in getting me to this point and I can only hope that I have made my gratitude and appreciation well-known to everyone that deserves it.

Thanks also to my many funding sources, including the Regents Fellowship, the NSF GRFP, Zobell fellowship, the CCE LTER program, and other SIO and UCSD graduate student funding opportunities.

Chapter 1, in full, is a reprint of the material as it appears in ISME 2018. Nagarkar, Maitreyi; Countway, Peter; Yoo, Yeong Du; Daniels, Emy; Poulton, Nicole; Palenik, Brian. 2016. The dissertation author was the primary author on this paper.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Nagarkar, Maitreyi; Wang, Maggie; Valencia, Bellineth; Palenik, Brian. The dissertation author was the primary author on this paper.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Nagarkar, Maitreyi; Palenik, Brian. The dissertation author was the primary author on this paper.

VITA

2011 Bachelor of Sciences, College of William & Mary

2016 Master of Sciences, Scripps Institution of Oceanography

2019 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Nagarkar M, Countway PD, Du Yoo Y, Daniels E, Poulton NJ, & Palenik B. (2018). Temporal dynamics of eukaryotic microbial diversity at a coastal Pacific site. The ISME journal 12: 2278–2291.

Gleason F, **Nagarkar M,** Chambouvet A, & Guillou L. A review of the characteristics of the dinoflagellate parasite Ichthyodinium chabelardi and its potential effect on fin fish populations. Marine & Freshwater Research. In press.

Giron-Nava A, James CC, Johnson AF, Dannecker D, Kolody B, Lee A… **Nagarkar M**… & Sugihara G. (2017). Quantitative argument for long-term ecological monitoring. Marine Ecology Progress Series 572:269-274.

Huyck RW, **Nagarkar M,** Olsen N, Clamons SE, & Saha MS. (2015). Methylmercury exposure during early *Xenopus laevis* development affects cell proliferation and death but not neural progenitor specification. Neurotoxicology and teratology 47:102-113.

Bauman SJ, Costa MT, Fong MB, House BM, Perez EM, Tan MH… **Nagarkar M**… & Franks PJS. (2014). Augmenting the biological pump: The shortcomings of geoengineered upwelling. Oceanography 27(3):17–23.

Lewis B, Wester MR, Miller LM, **Nagarkar MD,** Johnson MB, & Saha MS. (2009). Cloning and Characterization of Voltage-Gated Calcium Channel Alpha1 Subunits in Xenopus laevis during Development. Developmental Dynamics 238:2891–2902.

ABSTRACT OF THE DISSERTATION

Temporal Dynamics of Marine Microbial Communities at the SIO Pier

by

Maitreyi Dnyanadeep Nagarkar

Doctor of Philosophy in Marine Biology

University of California San Diego, 2019

Professor Brian Palenik, Chair

Marine microbial communities consist of millions of species engaging in complex interactions with one another and with their environments over a variety of time scales. The field of marine microbial ecology has only begun to understand the true extent of the diversity – both of species themselves (which include members of the bacteria, archaea, and eukaryotic protists) and the ways they interact with one another. Many studies of marine species diversity represent snapshots of a community but do not capture the temporal dynamics of its members. In this dissertation I illustrate how the microbial community at a single site changes over time by collecting high-frequency samples at the Scripps Pier (La Jolla, CA, USA). I leverage amplicon

sequencing to describe the bacterial and eukaryotic communities and find there to be detectable, occasionally very large, fluctuations, some with seasonal patterns and others on the order of days. I use this high-resolution sequence data to: identify putative grazers of the marine cyanobacteria *Synechococcus* (Chapter 1), characterize fine-scale changes in *Synechococcus* microdiversity over the course of blooms (Chapter 2), and suggest new Syndiniales parasite-host interactions at our site (Chapter 3).

INTRODUCTION

Microbial communities form the foundation of marine ecosystems. A single bucket of seawater contains minimally thousands of ecological interactions, from trophic (predation, parasitism) to competitive (allelopathy, nutrient scavenging, signal recognition), typically occurring on time-scales ranging from seconds to weeks. Identifying and characterizing these interactions, and the organisms involved in them, informs our understanding of concepts like: how energy and carbon move through the marine food web; how stable and resilient a community is over time, the effect of environmental factors on ocean life, and even the effects of marine microbial communities on human health. Given the incredible complexity of these ecosystem dynamics, and the great diversity of microbial species involved, there are many challenges to identifying and validating microbial ecological processes. But at a time when the oceans are changing in an unprecedented manner, it is very important to understand all aspects of marine ecosystems. In this dissertation I describe an extensive study of the microbial community at a single site, the research pier at the Scripps Institution of Oceanography. Through frequent (weekly or bi-weekly) sampling and high-throughput sequencing techniques, I have been able to describe the diversity of the SIO pier's microbial community as it changes with time. In addition to gaining a dynamic picture of the community, I have used this large dataset to ask questions about different types of ecological interactions at our site.

The challenge of describing microbial diversity

The important role of microbial activities in marine ecosystems has been increasingly recognized in the past few decades. The photosynthetic marine cyanobacteria, discovered in

1979, are responsible for up to 40% of marine carbon fixation (Agawin et al., 2000; Waterbury et al., 1979). Heterotrophic bacteria recycle particulate organic matter when they consume sinking particles, allowing these nutrients and carbon to once again become available to primary producers (Azam et al., 1983). Heterotrophic and mixotrophic unicellular eukaryotes ("protists") are the primary consumers in marine ecosystems and comprise a crucial but once-overlooked trophic step in marine food webs (Fenchel, 1982). Microbial parasites, including the ubiquitous and abundant members of the alveolate group Syndiniales, are an under-characterized source of mortality for a variety of species. Each new discovery of important microbial groups has motivated major modifications to our conception of marine food webs and the functional capabilities of marine microbial communities, expanding these dynamics beyond simple, stepwise trophic interactions.

The earliest recorded studies of microorganisms, including possibly the first marine microorganisms, were observations made using compound microscopes by 17th century scientists Robert Hooke and Antony van Leeuwenhoek. The term "protist" was likely coined by Ernst Haeckel, who completed thousands of detailed drawings of marine organisms, including phytoplankton such as radiolarians, foraminiferans, and diatoms (Karl & Proctor, 2007). In the 19th and 20th century marine microbiology shifted from being purely observational to including laboratory- or field-based experimental work. This included cultivation and maintenance of marine microbial species by researchers such as Claude ZoBell (Karl & Proctor, 2007). It was soon observed that only a small fraction of bacteria that could be observed under a microscope were able to be cultivated in a laboratory setting. This phenomenon, sometimes called the "Great Plate Count Anomaly," was an important indicator that alternative methods would be necessary to document the true diversity of marine microbial life (Staley, James T; Konopka, 1985).

In the 1980s Carl Woese proposed the use of sequencing DNA encoding ribosomal subunits, starting with 16S in the bacteria, to more fully characterize microbial communities (Woese et al., 1975, 1977). These rRNA-coding regions include segments of both highly conserved regions and hypervariable regions that are presumably under neutral selection, making them perfect candidates for distinguishing among different species; theoretically, primers could bind to every organism with a conserved region, but the sequences in the hypervariable regions of different organisms would differ. The distance between two sequences would be indicative of their evolutionary distance (Woese et al., 1977). This technique could be applied to broad groups of taxa (like describing the entire eukaryotic community using the 18S rRNA region) or to primers that target a more specific range of organisms (like a primer specific to *Synechococcus* for an ITS region that is able to provide information distinguishing between different *Synechococcus* clades).

The adoption of this technique transformed our ability to describe microbial communities and the advent of high-throughput sequencing allowed it to be applied at an unprecedented order of magnitude. Sequencing of the 18S gene has revealed far greater species diversity among microbial eukaryotes compared to traditional microscopy, including information motivating the restructuring of known phylogenies as well as the discovery of entirely novel clades (Epstein & López-García, 2008; Massana et al., 2014; Moon-van der Staay et al., 2001; Seenivasan et al., 2013). A significant proportion of this diversity has still not been classified or characterized, as many sequences represent organisms that have never been observed. There has recently been a strong push to sequence microbiomes globally, be they host-associated, terrestrial, or marine, and these have uncovered at least tenfold greater microbial diversity than previously described (Lima-mendez et al., 2015; Not et al., 2009; Rusch et al., 2007; Thompson et al., 2017; de

Vargas et al., 2015).

Characterizing marine microbial communities is the first step to exploring ecological interactions and can now be done in unprecedented detail. In this dissertation I use environmental sequencing to describe the overall microbial community at the Scripps Pier and to provide information about three different types of ecological interactions: grazing, competition, and parasitism. I focus specifically on protists, or unicellular eukaryotic microbes, and the cyanobacterial genus *Synechococcus*, a globally important primary producer.

Grazing in marine food webs

Predation on the microbial level is often referred to as "grazing," wherein unicellular zooplankton are the primary consumers of phytoplankton (including cyanobacteria). Azam et al. (1983) noted that a significant portion of water column bacteria were consumed by nano- and microflagellates, which not only indicated that marine food webs had previously undescribed trophic steps, but also led to the hypothesis that this contributes to release of fixed carbon back into the water as DOM (the "microbial loop"). The contribution of grazing to phytoplankton mortality has been measured in many parts of the ocean (Neuer & Cowles, 1994; Pasulka et al., 2015; Selph et al., 2011) and the role of small grazers in any system directly impacts the number of trophic links and thus efficiency of carbon transfer (Landry & Calbet, 2004; Sherr & Sherr, 1988).

Grazing of picophytoplankton has long been thought to be governed by cell-size and predator-prey contact rates (Gonzalez, 2004). However, more recent evidence has implicated other factors in driving grazer selectivity, with the quality of the food source playing a role in whether the prey is taken up or digested. *Synechococcus* has generally been considered a poor-

quality food source for grazers, and other studies have demonstrated similar preferential grazing strategies that are not accounted for by cell size, elemental ratios, or motility, indicating a need for further study of potential resistance mechanisms and other factors driving these predator-prey dynamics (Apple et al, 2011). While *Synechococcus* grazers have been observed and identified in many studies (Christaki et al., 1999; Frias-Lopez et al., 2009; Tsai et al., 2007), much of the work on grazing has accounted for the grazers themselves as a whole group, and individual impacts of different grazers remain to be well characterized.

Microdiversity in cyanobacteria

The incredible amount of diversity discovered using high-throughput sequencing has further complicated Hutchinson's "Paradox of the Plankton," the still-open question of how so many species that seemingly occupy the same niche can co-occur in the ocean (Hutchinson, 1961). This is an underlying theme of any marine microbiology study, as refining our knowledge of diversity in light of environmental conditions or species-species interactions can move us closer to understanding the more individual niches that allow so many organisms to thrive. The cyanobacterial genera *Synechococcus* and *Prochlorococcus* are already known to comprise multiple strongly-supported clades, some of which are well associated with light, temperature, salinity, or nutrient regimes (Ahlgren & Rocap, 2012; Flombaum et al., 2013; Rocap et al., 2002; Scanlan et al., 2009; Sohm et al., 2016). Yet, it is still not entirely understood why certain clades can consistently co-occur, as do *Synechococcus* clades I and IV at coastal sites (Tai & Palenik, 2009).

Additionally, recent work has uncovered diversity at even finer resolutions. Sequencing of seawater samples has revealed a wide diversity of co-existing variants within the same

cyanobacterial species or clade and even variants of small amplicons have been associated with differing genomic content (Kashtan et al., 2014; Lee et al., 2019). Studies have also found temporal structuring of dynamics between closely-related *Synechococcus* variants (Mackey et al., 2017). At the SIO pier and in the southern California Current Ecosystem the presence of different variants has been recognized and some functional differences between variants are known (Stuart et al., 2009; Tai et al., 2011). However, the abiotic and biotic factors that might contribute to the relative fitness of two very closely related variants are not fully understood and could inform our understanding of within-species competition.

Our emerging understanding of parasitism in the ocean

Parasitism has long been undervalued in its ecological significance. Parasites are actors at every trophic level and in some ways function predators (or grazers), but tend to have more complicated life cycles with different stages that must be considered separately in terms of their ecological roles (Anderson & May, 1978). When parasitic interactions are incorporated into food webs they often become dominant interaction type and increase the connectance of the system (Lafferty et al., 2006, 2008). Parasitic interactions could be an important driver of population dynamics to consider in marine microbial systems; for example, some studies have hypothesized that certain dinoflagellate blooms occur only when the dinoflagellate species is able to escape the pressure of parasitism, and that the subsequent termination of those blooms is a result of mortality due to parasitic infections (Chambouvet et al., 2008; Montagnes et al., 2008).

In the ocean, known parasitic species are members of numerous planktonic taxa including the chytrids, cercozoans, apicomplexans, perkinsozoans and Syndiniales, with planktonic hosts

of all sizes, ranging from bacteria to copepods (Skovgaard, 2014). The Syndiniales parasites in particular were observed in the 1950s when Hollande and Cachon (1952) noticed what is now called *Ichthyodinium chabelardi* in the eggs and larvae of fishes and carefully documented its life cycle, including a free-living, infective stage. Thereafter other parasitic Syndinians were observed with a wide variety of hosts, including copepods (Skovgaard et al., 2005), ciliates (Jung et al., 2016), crustaceans (Stentiford & Shields, 2005), and many different dinoflagellates (Brosnahan et al., 2014; Coats et al., 2002; Guillou et al., 2008; Kim & Park, 2014; Park et al., 2004; Siano et al., 2011). But their high representation in 18S sequencing efforts (Cleary & Durbin, 2016; de Vargas et al., 2015) has underscored the need for further study, especially given that several many well-supported Syndiniales clades (as designated based on environmental 18S sequences) have never been observed or maintained in a lab.

Limitations of the environmental sequencing approach

As powerful as environmental sequencing is for uncovering new diversity and describing communities in all their complexity, there are many important challenges and limitations to consider. Firstly, for amplicon sequencing to accurately identify the species present in an environmental sample, the region of interest must be present in all organisms that the study wishes to describe. Furthermore, the primers used to amplify the region must be free from bias. If these requirements are not met than entire groups of taxa may be nearly or entirely absent from the resulting sequence data, as was found with a commonly used set of 16S primers that greatly underrepresented the abundant bacterial group SAR11 (Parada et al., 2016). Primer choice has been found to have an effect in numerous other studies as determined by in silico studies (Bradley et al., 2016; Hadziavdic et al., 2014; Hugerth et al., 2014), comparisons of different

primer sets for the same region (Hugerth et al., 2014), and comparisons of sequencing results to known mock communities (Bradley et al., 2016; Fouhy et al., 2016; Geisen et al., 2015; Pinto & Raskin, 2012). Many other methodological issues, including the ubiquity of contamination in sequencing plates (Lusk, 2014), have also been identified as sources of error in this type of work.

Another important consideration in the use of amplicon sequence data is the fact that the number of sequence reads is not representative of the abundance of an organism within a sample or between samples. The former could be due to primer biases in amplification as mentioned above, or because of other issues like copy number variation between different organisms (Zhu et al., 2005). This leads, for example, to the well-described phenomenon of dinoflagellates being highly over-represented in sequencing studies, very likely because of their high copy numbers of the 18S region. Thus a dinoflagellate species could dominate the sequence data even if there are actually fewer cells of it than other species present. Additionally, amplicon sequence data is compositional in nature. Standard methods of sequencing environmental samples do not offer information about the actual quantity of biomass in the sample, and abundances of taxa found in a single sequenced sample are actually *relative* abundances. An artifact of this compositionality is that the sum of all sequences within a sample must equal one, so the relative abundance is heavily dependent on how many other species are present, and between samples the relative abundance of a taxon can decrease even if its cell number actually increases (Friedman & Alm, 2012; Gloor et al., 2016). Numerous methods have been proposed to address the issue of compositionality, both analytical (Friedman & Alm, 2012; Kurtz et al., 2015) and methodological (Satinsky et al., 2013; Ushio et al., 2018; Wang et al., 2018).

Finally, it is important to recognize that our ability to use sequence data to assign classifications to the members of a microbial community is only as good as how informative the

amplicon is (based on length and position) and how much resolution the available taxonomic databases can provide. Additionally, numerous methods are currently in use for assigning amplicon sequences to taxonomic units, including simply clustering based on a sequence similarity threshold (Schloss et al., 2009) and identifying 'true' sequences based on error and entropy models (Amir et al., 2017; Callahan et al., 2016).

All of these findings indicate that there is much room for improvement in the way that microbial community sequence data is collected and analyzed. Throughout this dissertation I have attempted to explore these limitations where possible, and use the results to inform how I interpret the data. For example, in Chapter 1 I examine whether a 97% clustering threshold is too loose and masks dynamics between what might be distinct taxa. In Chapters 2 and 3 I sequence mock communities alongside my environmental samples to get a sense of how well the sequencing results capture the known richness and evenness (i.e. the number of taxa present and their relative abundances) of the mock communities. And in Chapter 3 I try two methods of adjusting the relative abundance data to make it semi-quantitative rather than compositional.

Outline of the dissertation

In this dissertation I take advantage of a high-frequency sampling dataset at a well-studied coastal site, the Scripps Pier, to investigate temporal dynamics of different types of microbial interactions.

In Chapter 1, I examine potential *Synechococcus* grazers by sequencing the eukaryotic community (using the 18S amplicon) during different time points of interest in the *Synechococcus* population (such as periods of rapid bloom onset and decline). Coupling the

sequencing results with some flow-sorted populations of putative grazer cells, I suggest two putative mixotrophic grazers of *Synechococcus*.

In Chapter 2, I sequence an amplicon that provides higher-resolution of *Synechococcus* variants at the sub-clade level to examine within-*Synechococcus* diversity temporally over the course of three different blooms, as well as spatially at different sites in the Southern California ecosystem. The results reveal there to be shifts in the relative abundances of both broad *Synechococcus* clades and higher-resolution variants within those clades over time and space.

In Chapter 3, I explore a taxonomic group that was abundant in the sequence data from chapter two: Syndiniales. This group consists of at least five clades of parasitic species and its sequences have been identified ubiquitously in studies around the world. I identify some putative Syndiniales-host interactions using time series analysis methods and a weekly or bi-weekly 18S sequence dataset.

The Scripps Pier site has been a marine monitoring site for over a century and offers a wealth of measurements over the years pertaining to biological and environmental variables. Having access to this site allows for regular, frequent sampling, which results in a time series dataset from a single location. While there has often been an emphasis on comparing diversity spatially, there is a strong need for more temporal studies, both to establish baselines and detect changes within ecosystems. In this dissertation, I leverage a high-frequency sampling dataset to describe in detail the temporal dynamics at the SIO pier.

REFERENCES

Agawin, N. S. R., Duarte, C. M., & Agustí, S. (2000). Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. *Limnology and Oceanography*, *45*(3), 591–600. https://doi.org/10.4319/lo.2000.45.3.0591

Ahlgren, N. A., & Rocap, G. (2012). Diversity and distribution of marine Synechococcus : multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Frontiers in Microbiology*, *3*(June), 1–24. https://doi.org/10.3389/fmicb.2012.00213

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., et al. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *MSystems*, *2*(2), 1–7.

Anderson, R. M., & May, R. M. (1978). Regulation and Stability of Host-Parasite Population Interactions. *Journal of Animal Ecology*, *47*(1), 219–247. https://doi.org/10.2307/3933

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. a, & Thingstad, F. (1983). The Ecological Role of Water-Column Microbes in the SEa. *Marine Ecology Progress Series*. https://doi.org/10.3354/meps010257

Bradley, I. M., Pinto, A., & Guest, J. S. (2016). Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities. *Applied and Environmental Microbiology*, *82*(19), 5878–5891. https://doi.org/10.1128/AEM.01630-16.Editor

Brosnahan, M. L., Farzan, S., Keafer, B. A., Sosik, H. M., Olson, R. J., & Anderson, D. M. (2014). Complexities of bloom dynamics in the toxic dinoflagellate Alexandrium fundyense revealed through DNA measurements by imaging flow cytometry coupled with species-specific rRNA probes. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *103*, 185–198. https://doi.org/10.1016/j.dsr2.2013.05.034

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Chambouvet, A., Morin, P., Marie, D., & Guillou, L. (2008). Control of Toxic Marine Dinoflagellate Blooms by Serial Parasitic Killers. *Science*, *322*(5905), 1254–1257. https://doi.org/10.1126/science.1164387

Christaki, U., Dolan, J. R., Vaulot, D., & Rassoulzadegan, F. (1999). Growth and grazing on Prochlorococcus and Synechococcus by two marine ciliates, *44*(1), 52–61.

Cleary, A. C., & Durbin, E. G. (2016). Unexpected prevalence of parasite 18S rDNA sequences

in winter among Antarctic marine protists. *Journal of Plankton Research*, *0*(0), 1–17. https://doi.org/10.1093/plankt/fbw005

Coats, D. W., Coats, D. W., Park, M. G., & Park, M. G. (2002). Parasitism of photosynthetic dinoflagellates by three strains of. *Aquatic Microbial Ecology*, *528*, 520–528. https://doi.org/10.1046/j.1529-8817.2002.01200.x

Epstein, S., & López-García, P. (2008). "Missing" protists: A molecular prospective. *Biodiversity and Conservation*, *17*(2), 261–276. https://doi.org/10.1007/s10531-007-9250-y

Fenchel, T. (1982). Ecology of Heterotrophic Microflagellates. IV Quantitative Occurrence and Importance as Bacterial Consumers . *Marine Ecology Progress Series*, *9*(1977), 35–42. https://doi.org/10.3354/meps009035

Flombaum, P., Gallegos, J. L., Gordillo, R. a, Rincón, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine Cyanobacteria Prochlrococcus and Synechococcus. *Pnas*, *110*(24), 9824–9829. https://doi.org/10.1073/pnas.1307701110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1307701110

Fouhy, F., Clooney, A. G., Stanton, C., Claesson, M. J., & Cotter, P. D. (2016). 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method , primer choice and sequencing platform. *BMC Microbiology*, 1–13. https://doi.org/10.1186/s12866-016-0738-z

Frias-Lopez, J., Thompson, A., Waldbauer, J., & Chisholm, S. W. (2009). Use of stable isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface waters. *Environmental Microbiology*, *11*(2), 512–525. https://doi.org/10.1111/j.1462-2920.2008.01793.x

Friedman, J., & Alm, E. J. (2012). Inferring Correlation Networks from Genomic Survey Data. *PLoS Computational Biology*, *8*(9), 1–11. https://doi.org/10.1371/journal.pcbi.1002687

Geisen, S., Laros, I., Vizcaíno, A., Bonkowski, M., & De Groot, G. A. (2015). Not all are free-living: High-throughput DNA metabarcoding reveals a diverse community of protists parasitizing soil metazoa. *Molecular Ecology*, *24*(17), 4556–4569. https://doi.org/10.1111/mec.13238

Gloor, G. B., Rong, J., Pawlowsky-glahn, V., & José, J. (2016). It's all relative : analyzing microbiome data as compositions. *Annals of Epidemiology*, *26*(5), 322–329. https://doi.org/10.1016/j.annepidem.2016.03.003

Gonzalez, J. M. (2004). Efficient size-selective bacterivory by phagotrophic nanoflagellates in aquatic ecosystems. *Marine Biology*, *126*(4), 785–789. https://doi.org/10.1007/bf00351345

Guillou, L., Viprey, M., Chambouvet, A., Welsh, R. M., Kirkham, A. R., Massana, R., et al. (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to

Syndiniales (Alveolata). *Environmental Microbiology*, *10*(12), 3349–3365. https://doi.org/10.1111/j.1462-2920.2008.01731.x

Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E. M., & Troedsson, C. (2014). Characterization of the 18s rRNA gene for designing universal eukaryote specific primers. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0087624

Hollande, A., & Cachon, J. (1952). Un parasite des oeufs de sardine: l'I chthyodinium chabelardi nov. gen. nov. sp. (Péridinien parasite). *C R Acad. Sci*, *235*, 976–977.

Hugerth, L. W., Muller, E. E. L., Hu, Y. O. O., Lebrun, L. A. M., Roume, H., Lundin, D., et al. (2014). Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. *PLoS ONE*, *9*(4). https://doi.org/10.1371/journal.pone.0095567

Hutchinson, G. E. (1961). The Paradox of the Phytoplankton. *The American Naturalist*, *95*(882), 137–145.

Jung, J. H., Choi, J. M., Coats, D. W., & Kim, Y. O. (2016). Euduboscquella costata n. Sp. (Dinoflagellata, Syndinea), an Intracellular Parasite of the Ciliate Schmidingerella arcuata: Morphology, Molecular Phylogeny, Life Cycle, Prevalence, and Infection Intensity. *Journal of Eukaryotic Microbiology*, *63*(1), 3–15. https://doi.org/10.1111/jeu.12231

Karl, D. M., & Proctor, L. M. (2007). Foundations of Microbial Oceanography. *Oceanography*, *20*(2), 16–27.

Kashtan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., et al. (2014). Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild Prochlorococcus, *344*(April), 416–421.

Kim, S., & Park, M. G. (2014). Amoebophrya spp. from the bloom-forming dinoflagellate cochlodinium polykrikoides: Parasites not nested in the "amoebophrya ceratii complex." *Journal of Eukaryotic Microbiology*, *61*(2), 173–181. https://doi.org/10.1111/jeu.12097

Kurtz, Z. D., Muller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLoS Computational Biology*, *11*(5), 1–25. https://doi.org/10.1371/journal.pcbi.1004226

Lafferty, K. D., Dobson, A. P., & Kuris, A. M. (2006). Parasites dominate food web links, *103*(30), 11211–11216. https://doi.org/10.1073/pnas.0604755103

Lafferty, K. D., Allesina, S., Arim, M., Briggs, C. J., De Leo, G., Dobson, A. P., et al. (2008). Parasites in food webs: the ultimate missing links. *Ecology Letters*, *11*(6), 533–46. https://doi.org/10.1111/j.1461-0248.2008.01174.x

Landry, M. R., & Calbet, A. (2004). Microzooplankton production in the oceans, *507*. https://doi.org/10.1016/j.icesjms.2004.03.011

Lee, M. D., Ahlgren, N. A., Kling, J. D., Walworth, N. G., Rocap, G., Saito, M. A., et al. (2019). Marine Synechococcus isolates representing globally abundant genomic lineages demonstrate a unique evolutionary path of genome reduction without a decrease in GC content. *Environmental Microbiology*, *00*, 1–10. https://doi.org/10.1111/1462-2920.14552

Lima-mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., et al. (2015). Lima-Mendez et al (2015) Determinants of community structure in the grobal plankton interactome.pdf, *348*(6237), 1–10.

Lusk, R. W. (2014). Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS ONE*, *9*(10), e110808. https://doi.org/10.1371/journal.pone.0110808

Mackey, K. R. M., Hunter-Cevera, K. R., Britten, G. L., Murphy, L. G., Sogin, M. L., & Huber, J. A. (2017). Seasonal Succession and Spatial Patterns of Synechococcus Microdiversity in a Salt Marsh Estuary Revealed through 16S rRNA Gene Oligotyping, *8*(August), 1–11. https://doi.org/10.3389/fmicb.2017.01496

Massana, R., del Campo, J., Sieracki, M. E., Audic, S., & Logares, R. (2014). Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *The ISME Journal*, *8*(4), 854–66. https://doi.org/10.1038/ismej.2013.204

Montagnes, D. J. S., Chambouvet, A., Guillou, L., & Fenton, A. (2008). Responsibility of microzooplankton and parasite pressure for the demise of toxic dinoflagellate blooms. *Aquatic Microbial Ecology*, *53*(2), 211–225. https://doi.org/10.3354/ame01245

Moon-van der Staay, S. Y., De Wachter, R., & Vaulot, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, *409*(6820), 607–610. https://doi.org/10.1038/35054541

Neuer, S., & Cowles, T. J. (1994). Protist herbivory in the Oregon upwelling system. *Marine Ecology Progress Series*, *113*(1–2), 147–162. https://doi.org/10.3354/meps113147

Not, F., del Campo, J., Balagué, V., de Vargas, C., & Massana, R. (2009). New insights into the diversity of marine picoeukaryotes. *PLoS ONE*, *4*(9). https://doi.org/10.1371/journal.pone.0007143

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters : assessing small subunit rRNA primers for marine microbiomes with mock communities , time series and global field samples, *18*, 1403–1414. https://doi.org/10.1111/1462-2920.13023

Park, M. G., Yih, W., & Coats, D. W. (2004). Parasites and phytoplankton, with special emphasis on dinoflagellate infections. *Journal of Eukaryotic Microbiology*, *51*(2), 145–155. https://doi.org/10.1111/j.1550-7408.2004.tb00539.x

Pasulka, A. L., Samo, T. Y. J., & Landry, M. R. (2015). Grazer and viral impacts on microbial growth and mortality in the southern California Current Ecosystem. *Journal of Plankton Research*, *37*(2), 320–336. https://doi.org/10.1093/plankt/fbv011

Pinto, A. J., & Raskin, L. (2012). PCR Biases Distort Bacterial and Archaeal Community Structure in Pyrosequencing Datasets, *7*(8). https://doi.org/10.1371/journal.pone.0043093

Rocap, G., Distel, D. L., Waterbury, J. B., & Chisholm, S. W. (2002). Resolution of Prochlorococcus and Synechococcus Ecotypes by Using 16S-23S Ribosomal DNA Internal Transcribed Spacer Sequences. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, *68*(3), 1180–1191. https://doi.org/10.1128/AEM.68.3.1180-1191.2002

Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., et al. (2007). The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biology*, *5*(3), 0398–0431. https://doi.org/10.1371/journal.pbio.0050077

Satinsky, B. M., Gifford, S. M., Crump, B. C., & Moran, M. A. (2013). *Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis*. *Microbial Metagenomics, Metatranscriptomics, and Metaproteomics* (1st ed., Vol. 531). Elsevier Inc. https://doi.org/10.1016/B978-0-12-407863-5.00012-5

Scanlan, D. J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W. R., et al. (2009). Ecological genomics of marine picocyanobacteria. *Microbiology and Molecular Biology Reviews : MMBR*, *73*(2), 249–99. https://doi.org/10.1128/MMBR.00035-08

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur : Open-Source , Platform-Independent , Community-Supported Software for Describing and Comparing Microbial Communities □. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Seenivasan, R., Sausen, N., Medlin, L. K., & Melkonian, M. (2013). Picomonas judraskeda Gen. Et Sp. Nov.: The First Identified Member of the Picozoa Phylum Nov., a Widespread Group of Picoeukaryotes, Formerly Known as "Picobiliphytes." *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0059565

Selph, K. E., Landry, M. R., Taylor, A. G., Yang, E., Measures, C. I., Yang, J., et al. (2011). Spatially-resolved taxon-specific phytoplankton production and grazing dynamics in relation to iron distributions in the Equatorial Pacific between 110 and 140 1 W. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *58*, 358–377. https://doi.org/10.1016/j.dsr2.2010.08.014

Sherr, E., & Sherr, B. (1988). Role of microbes in pelagic food webs: A revised concept. *Limnology and Oceanography*, *33*(5), 1225–1227. https://doi.org/10.4319/lo.1988.33.5.1225

Siano, R., Alves-De-Souza, C., Foulon, E., M. Bendif, E., Simon, N., Guillou, L., & Not, F. (2011). Distribution and host diversity of Amoebophryidae parasites across oligotrophic waters of the Mediterranean Sea. *Biogeosciences*, *8*(2), 267–278. https://doi.org/10.5194/bg-8-267-2011

Skovgaard, A. (2014). Dirty tricks in the plankton: Diversity and role of marine parasitic protists. *Acta Protozoologica*, *53*(1), 51–62. https://doi.org/10.4467/16890027AP.14.006.1443

Skovgaard, A., Massana, R., Balague, V., & Saiz, E. (2005). Phylogenetic position of the copepod-infesting parasite Syndinium turbo (Dinoflagellata, Syndinea). *Protist*, *156*(4), 413–423. https://doi.org/10.1016/j.protis.2005.08.002

Sohm, J. A., Ahlgren, N. A., Thomson, Z. J., Williams, C., Moffett, J. W., Saito, M. A., et al. (2016). Co-occurring Synechococcus ecotypes occupy four major oceanic regimes defined by temperature , macronutrients and iron, *10*(2), 333–345. https://doi.org/10.1038/ismej.2015.115

Staley, James T; Konopka, A. (1985). Microorganisms in Aquatic and Terrestrial Habitats. *Annual Review of Microbiology*, *39*, 321–346.

Stentiford, G. D., & Shields, J. D. (2005). A review of the parasitic dinoflagellates Hematodinium species and Hematodinium-like infections in marine crustaceans. *Diseases of Aquatic Organisms*, *66*(1), 47–70. https://doi.org/10.3354/dao066047

Stuart, R. K., Dupont, C. L., Johnson, D. A., Paulsen, I. T., Palenik, B., & Icrobiol, A. P. P. L. E. N. M. (2009). Coastal Strains of Marine Synechococcus Species Exhibit Increased Tolerance to Copper Shock and a Distinctive Transcriptional Response Relative to Those of Open-Ocean Strains □ †, *75*(15), 5047–5057. https://doi.org/10.1128/AEM.00271-09

Tai, V., & Palenik, B. (2009). Temporal variation of Synechococcus clades at a coastal Pacific Ocean monitoring site. *The ISME Journal*, *3*(8), 903–915. https://doi.org/10.1038/ismej.2009.35

Tai, V., Burton, R. S., & Palenik, B. (2011). Temporal and spatial distributions of marine Synechococcus in the Southern California Bight assessed by hybridization to bead-arrays. *Marine Ecology Progress Series*, *426*(MARCH), 133–147. https://doi.org/10.3354/meps09030

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., et al. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, *551*(7681), 457–463. https://doi.org/10.1038/nature24621

Tsai, A. Y., Chiang, K. P., Chan, Y. F., Lin, Y. C., & Chang, J. (2007). Pigmented nanoflagellates in the coastal western subtropical Pacific are important grazers on Synechococcus populations. *Journal of Plankton Research*, *29*(1), 71–77. https://doi.org/10.1093/plankt/fbl058

Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., et al. (2018). Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing, 1–15. https://doi.org/10.3897/mbmg.2.23297

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science (New York, N.Y.)*, *348*(6237), 1261605. https://doi.org/10.1126/science.1261605

Wang, S., Lin, Y., Gifford, S., Eveleth, R., & Cassar, N. (2018). Linking patterns of net community production and marine microbial community structure in the western North Atlantic. *The ISME Journal*, 2582–2595. https://doi.org/10.1038/s41396-018-0163-4

Waterbury, J. B., Watson, S. W., Guillard, R. R. L., & Brand, L. E. (1979). Widespread occurance of a unicellular, marine, planktonic, cyanobacterium. *Nature*.

Woese, C. R., George, E. F., Zablen, Lawrence Uchida, T., Bonen, L., Pechmen, K., J. Lewis, B., & Stahl, D. (1975). Conservation of primary structure in 16S ribosomal RNA. *Nature*, *254*, 83–86.

Woese, C. R., Fox, G. E., & Pechman, K. R. (1977). Comparative Cataloging of 16S Ribosomal Ribonucleic Acid: Molecular Approach to Procaryotic Systematics. *International Journal of Systematic and Evolutionary Microbiology*, *27*(1), 44–57. https://doi.org/10.1099/00207713-27-1-44

Zhu, F., Massana, R., Not, F., Marie, D., & Vaulot, D. (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiology Ecology*, *52*(1), 79–92. https://doi.org/10.1016/j.femsec.2004.10.006

CHAPTER 1:

Temporal dynamics of eukaryotic microbial diversity at a coastal Pacific site

1.1 ABSTRACT

High-throughput sequencing of ocean biomes has revealed vast eukaryotic microbial diversity, a significant proportion of which remains uncharacterized. Here we use a temporal approach to understanding eukaryotic diversity at the Scripps Pier, La Jolla, California, USA, via high-throughput amplicon sequencing of the 18S rRNA gene, the abundances of both *Synechococcus* and *Synechococcus* grazers, and traditional oceanographic parameters. We also exploit our ability to track OTUs temporally to evaluate the ability of 18S sequence-based OTU assignments to meaningfully reflect ecological dynamics. The eukaryotic community is highly dynamic in terms of both species richness and composition, though proportional representation of higher-order taxa remains fairly consistent over time. *Synechococcus* abundance fluctuates throughout the year. Operational Taxonomic Units (OTUs) unique to dates of *Synechococcus* blooms and crashes or enriched in *Synechococcus* addition incubation experiments suggest that the prasinophyte *Tetraselmis* sp. and *Gymnodinium*-like dinoflagellates are likely *Synechococcus* grazers under certain conditions, and may play an important role in their population fluctuations.

1.2 INTRODUCTION

Microbial communities form the foundation of marine ecosystems. Phytoplankton, including the cyanobacteria, are responsible for the majority of marine primary productivity (Pomeroy, 1974; Waterbury et al, 1979). Heterotrophic and mixotrophic eukaryotes, commonly known as grazers, are primary consumers that comprise a crucial but once-overlooked trophic step (Fenchel, 1982). Understanding basal trophic interactions relies on characterizing the long-underestimated diversity of marine microbial species. Two primary challenges have been the limitations of morphological species identification and the inability to identify most microbial

species (Potter et al, 1997). Environmental sequencing offers a potential solution: sequencing the 18S rRNA gene has revealed far greater species diversity among protists than that recorded using microscopy, including entirely novel clades (Moon-van der Staay et al, 2001; Massana et al, 2014; Seenivasan et al, 2013; Cheung et al, 2010; de Vargas et al, 2015; Not et al, 2008).

Many sequencing-based studies of microbial diversity focus on spatial comparisons, implicitly assuming that a single time-point can represent the community at a sampling station. However, temporal studies have revealed considerable variability in community composition, especially at coastal sites (Countway et al, 2010; Cleary & Durbin, 2016; Collado-Fabbri et al, 2011; Massana et al, 2004). While some taxa remain consistent in their abundances, others exhibit sudden, transient blooms (Goericke, 2011). These temporal dynamics are not fully understood, but may reveal information about trophic interactions as well as factors underlying harmful blooms.

Cyanobacteria, including members of the genus *Synechococcus,* play a prominent role in ocean primary productivity: they are estimated to comprise 8% of worldwide phytoplankton biomass and are responsible for up to 40% of marine carbon fixation (Waterbury et al, 1979; Agawin et al, 2000; Flombaum et al, 2013). *Synechococcus* species are globally abundant, especially in coastal environments (Olson et al, 1990; Partensky et al, 1999) such as the California Current Ecosystem (Tai et al, 2009).

The Scripps Institution of Oceanography pier, located in the Southern California Bight, has served as a long-term monitoring site since the early 1900s, when W.E. Allen recorded diatom and dinoflagellate abundances for over ten years (Allen, 1936). More recently, *Synechococcus* abundance has been monitored weekly or bi-weekly for at least twenty years (Tai et al, 2009; Palenik lab unpubl data), in addition to counts of harmful algal species and other

physical and chemical metadata (SCCOOS, http://www.sccoos.org/). These historical data make the Scripps Pier an excellent site for a temporal study of the microbial community.

We have utilized environmental sequencing to describe the composition of the eukaryotic community at the Scripps Pier, and to explore the temporal dynamics of individual protist species and their potential roles as drivers of the onset or termination of *Synechococcus* blooms. Seasonality in *Synechococcus* cell density has previously been observed at the Scripps Pier and other coastal sites (Tai et al, 2009; Moisan et al, 2010; Taylor et al, 2014; Agawin et al, 1998; Worden et al, 2004). The *Synechococcus* population at the Scripps Pier has historically bloomed once or twice every spring, typically followed by rapid declines of the population (days to weeks) back to basal levels (Palenik 2000; Tai et al, 2009; Palenik lab unpubl. data). The major processes responsible for loss of *Synechococcus* biomass are sedimentation and sinking via attachment to larger particles, viral lysis, and grazing—often by ciliates and nanoflagellates (Christaki et al, 2002). While some *Synechococcus* grazers have been isolated, including *Paraphysmonas* sp., *Pteridomonas* sp., *Gymnodinium* sp., and various ciliates (Apple et al, 2011; Zwirglmaier et al, 2009; Jeong et al, 2005; Paz-Yepes, manuscript in preparation), most have not been validated as significant players in the ecology of *Synechococcus*. It is therefore largely unknown whether the same grazers that are readily maintained in culture are also important grazers of *Synechococcus* in the ocean.

The objectives of this study were to utilize the large amount of information provided by a temporal dataset of 18S rRNA amplicon sequences to (i) describe the composition and dynamics of the eukaryotic microbial community at the Scripps Pier site; (ii) identify putative grazers of *Synechococcus*; and (iii) evaluate the ecological significance of cultured species of grazers and

other protists with respect to their role in shaping *Synechococcus* population dynamics at this site.

1.3 METHODS

1.3.1 Sample collection

Sampling was conducted in 2011 and 2012 as described in Tai et al (2009). Briefly, surface seawater was collected weekly from the end of the Scripps Pier in La Jolla, CA (32º87'N, 117º26'W). Sampling dates were first selected for amplicon sequencing based on their correspondence with *Synechococcus* or grazer blooms or declines, with additional time points included to increase overall temporal coverage. The resulting amplicon sequence data described seventeen time points covering diverse seasons over two years (Table 1.1).

1.3.2 DNA Extraction and Sequencing

For DNA extraction, 500 ml of seawater was filtered in triplicate onto 0.2 µm Supor filter discs (Pall) and stored at -80 ºC. DNA was extracted with the DNEasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA) and quantified by PicoGreen fluorescence (Quant-iT; Qiagen) as described in Tai et al (2009). DNA from the seventeen time-points was sent to Research & Testing Laboratories (Lubbock, TX, USA) for sequencing of the V4 region of the 18S gene on a Roche 454-FLX using the TAReuk454FWD1 (5'-CCAGCA(G/C)C(C/T)GCGGTAATTCC-3') and TAReukREV3 (5'-ACTTTCGTTCTTGAT(C/T)(A/G)A-3′) primers (Stoeck et al, 2010). For PCR conditions and complete methods, see S1.

1.3.3 Sequence Analysis

Sequence reads were quality-filtered and trimmed to 245 bp using the mothur software (version 1.35.1) with the pipeline described in the mothur 454 SOP (Schloss et al 2009 & 2011,

accessed 15 Sept 2015). Chimeras were removed using UCHIME (Edgar et al, 2011) and the remaining sequences from each time-point were merged together for Operational Taxonomic Unit (OTU) assignment. Sequences were clustered into OTUs at different similarity cutoffs ranging from 100% to 95%. Representatives from each OTU were classified based on the SILVA v123 taxonomy (Quast et al, 2012). Sequences from each time-point were sub-sampled to the smallest sample size (1308 sequences) to prevent over-representing rare OTUs in more deeply sequenced samples. Certain sequences of interest were aligned on BLAST or SINA (Pruesse et al, 2012) for further taxonomic specificity. Unweighted Bray-Curtis analysis was conducted using SciPy (Jones et al, 2001). The vegan package in R (Oksanen et al, 2013; R Core team) was used to conduct a canonical correspondence analysis of sample variation as explained by five variables: chlorophyll, alpha diversity, *Synechococcus* density, grazer density, and water temperature.

1.3.4 Data Availability

Raw sequence reads and metadata were deposited into the NCBI Sequence Read Archive under accession number SRP132203 (https://www.ncbi.nlm.nih.gov/sra/SRP132203). They are also available on the BCO-DMO page (https://www.bco-dmo.org).

1.3.5 Flow Cytometry

Four replicate 1-ml seawater samples were preserved at a final concentration of 0.25% glutaraldehyde (Sigma-Aldrich, St Louis, MO, USA) for flow cytometry; upon addition of glutaraldehyde, samples were kept at room temperature for ten minutes and then stored at -80 ^{o}C. Samples were processed as described in Collier et al (2003) (S2). Cell abundance counts were normalized using pre-counted 0.94 µm green fluorescent beads (Duke Scientific Corporation, Palo Alto, CA, USA) that were added to each sample. Events falling within a specific region

were counted as *Synechococcus* cells (Fig. S3). Three-ml samples were frozen at -80 °C in 6.8% glycine betaine and sent to Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, for bulk flow-sorting. Sorting gates were established for two different regions of the flow-cytogram plots based on fluorescence, cell-size, and internal cell-complexity (scatter). The objective was to capture putative mixotrophic grazers, which we expected to be larger in size than *Synechococcus* and emit both chlorophyll fluorescence (from having chloroplasts) and phycoerythrin fluorescence (from ingested *Synechococcus*). Bulk sorts of two specific regions of interest (R1 and R2, Fig. 4D) containing putative mixotrophic grazers were completed as described in S2, collected in UV-treated sterile 0.5 ml centrifuge tubes and stored immediately at -80 °C.

1.3.6 Clone Libraries

Whole genome amplifications were performed on cells from R1 and R2 (Fig. 1.5D). Cells were lysed in 3 rounds of heating (70°C, 3 min)/freezing (-80°C, 3 min) with vortex bead-beating (1 minute, using ~50, 0.5 mm silica-zirconia beads per tube) in between the heating and freezing steps. The resulting crude cell-lysate was used in 20 µl Genomiphi V3 (GE Life Sciences, Marlborough, MA, USA) reactions that were primed with random hexamers. Amplification of DNA was confirmed by quantification on a Qubit fluorometer using the Broad Range Qubit Reagents (Thermo Fisher Scientific).

The clone libraries for R1 and R2 were made slightly differently but ultimately resulted in sequence data from the same region. For R1, a larger region of the 18S rRNA gene was initially amplified using the GoTaq HotStart Mastermix (Promega) with the Moon-A (5'-ACCTGGTTGATCCTGCCAG-3') and Moon-B (5'-TGATCCTTCYGCAGGTTCAC-3') primers (Moon-Van der Staay et al, 2001); the reaction was denatured at 94 °C for 2 minutes, then underwent 35 cycles of 30 seconds at 94 °C, 2 minutes at 52 °C, and 3 minutes at 72 °C,

followed by a final extension for 10 minutes at 72 °C. The PCR product was ligated into pCR2.1- TOPO (Thermo Fisher Scientific) plasmid vectors and transformed into TOP10 chemically competent *E. coli* cells following the manufacturer's protocol (Invitrogen, Carlsbad, CA, USA). Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen). The TAReuk454FWD1 and TAReukREV3 primers (Stoeck et al, 2010) were used as internal primers for Sanger-based DNA sequencing (Eton Biosciences, San Diego, CA, USA) of the cloned DNA templates to sequence the same V4 region of the 18S rRNA gene as the high-throughput sequencing. For R2, since the Moon A- Moon B amplification was deemed unnecessary, the initial amplification was conducted directly using the TAReuk primers (using the procedure from Stoeck et al 2010 and an annealing temperature of 45 °C for the last 25 cycles). PCR products were cloned as described for R1 and sequenced from the flanking M13F and M13R priming sites on the pCR2.1-TOPO vector.

1.3.7 Microscopy

On every sampling date, two 20 ml replicate seawater samples were preserved with glutaraldehyde (0.5%, final concentration), stained with DAPI and vacuum-filtered onto 2 µm polycarbonate filters (Spectrum). Filters were applied to a glass microscope slide with Immersol immersion oil (Zeiss, Oberkochen, Germany) and stored in the dark at -20 °C. Twenty pre-programmed random automated positions were counted on each slide using a Zeiss Axiovert 200 M inverted compound microscope. Grazers were identified as cells that demonstrated DAPI fluorescence within a defined nuclear region, a clear cell shape under FITC filter settings, and localized phycoerythrin fluorescence (blue light, 488 nm) suggesting consumption of one or more *Synechococcus* cells.

1.3.8 Synechococcus Enrichment Experiments

On 29 July 2014, six flasks of 600 ml freshly collected seawater were placed in a rooftop incubator filled with running seawater at ambient temperature. The flasks were kept below the surface of the water using ring-weights and placed under shaded mesh to prevent photo-degradation of cells. Cultured representatives from four *Synechococcus* clades (clade I: CC9311, clade II: CC9605, clade III: WH9102, clade IV: CC9902) were spun down, re-suspended in F/4 culture media (Guillard, 1975), and added to the seawater at a final concentration of $1 * 10^6$ cells ml^{-1}. After five days, 250 ml of seawater from each of these incubations was filtered, extracted for DNA, and 454-sequenced as described above.

1.3.9 *Synechococcus feeding experiments*

A *Tetraselmis* strain was isolated and cultured from a surface water sample collected at the Scripps Pier in January 2013. Mixtures of this *Tetraselmis* and *Synechococcus* cells were established and observed for ingestion and calculation of ingestion rates. More information on these feeding experiments is provided in S4.

1.4 RESULTS

1.4.1 Eukaryotic community diversity

Eukaryotic diversity was assessed at seventeen time-points during a two-year span (Table 1). A total of 67,295 DNA sequences were obtained after quality-trimming and alignment. Although multiple approaches are possible for interpreting the ecological implications of environmental sequence data (Eren et al, 2013; Mahé et al, 2014), we employed average-neighbor clustering methods to analyze the taxa (OTUs) in our samples. The seventeen samples together contained 6945 OTUs (before rarefaction) when using a 97% sequence identity cutoff to

assign approximate ‘species-level’ OTUs. Hereafter, “OTUs” will refer to those assigned at 97% cutoff unless otherwise specified, and the number in the name of an OTU reflects its overall abundance (OTU 1 refers to the most abundant OTU and so on). See Supplementary Table (S8). Rarefaction curves of each sample indicate that few samples were ‘saturated’ at the level of sub-sampling, suggesting that greater sequencing depth would be required to represent the entire microbial community for most dates (Fig. 1.1A). In total, approximately 0.6% of OTUs were ‘abundant’ (comprising 1% or more of total sequences), while 68% were ‘rare’ (fewer than 0.01% of total sequences). A wide range of taxonomic diversity was still present in the rare sequences, though a higher proportion were ‘unclassified’ (Fig. S5). After sub-sampling to the least common number of sequences (n=1308) at each time point the total number of OTUs across all time-points was 2110. The estimated species richness on each date (using the Chao1 indicator) was highly variable, ranging from 80 to 961 OTUs (Fig. 1.1B).

Relating sequences to eukaryotic taxa requires classification of OTUs; we did this using mothur (Schloss et al, 2009) and the Silva v123 database (Quast et al, 2012). The eukaryotic community composition was temporally dynamic, with alveolates consistently comprising the largest proportion of the classified sequences (Fig. 1.2A). Alveolates are known to have high copy numbers of the 18S amplicon (Prokopowich et al, 2003) so it is unclear whether their prominence in environmental sequence data can be related to their abundance. We therefore focus on changes in the relative (rather than the absolute) abundance of OTUs in our data. While representatives from all other major protistan lineages were present at nearly every time-point, their relative abundances varied throughout the sampling period. A large portion of the sequences and OTUs were unclassified, even at the highest taxonomic levels, suggesting a large amount of novel eukaryotic diversity. The sampling dates that were most similar based upon five

variables (chlorophyll, Chao1 diversity, temperature, grazer density, and *Synechococcus* density), as determined by canonical correspondence analysis (Fig. 1.1C), did not necessarily cluster together based on the similarities of their eukaryotic communities (Fig. 1.1D). The microbial eukaryotic communities that were sampled during *Synechococcus* blooms and peaks in grazer abundance were slightly more similar to one another than they were to the communities at other time-points, as demonstrated by Bray-Curtis dissimilarity (Fig. 1.1D).

Alveolates also dominated the phylogenetic diversity, represented as the proportion of total OTUs (not sequences) within each taxonomic group (Fig. 1.2B). Some other groups had a relatively low abundance of overall sequence numbers, but high phylogenetic diversity within the group. For example, on average picomonads were represented by 3% of total OTUs but only 0.2% of the total number of sequences. The relative richness of OTUs (number of OTUs per higher-level taxonomic group) was temporally stable compared to the proportional representation of sequence-types to the total sequence count. For example, while the relative proportion of alveolate sequences ranged widely between time-points (Fig. 1.2A), the number of alveolate OTUs was consistently around a quarter of total OTUs (Fig. 1.2B). The most abundant alveolates included dinoflagellates that have been previously found be ecologically important at the Scripps Pier, including *Prorocentrum, Lingulodinium,* and *Akashiwo* (SCCOOS, http://www.sccoos.org/; Goericke 2011).

Individual OTUs varied greatly in their temporal distributions. Of the twenty most abundant OTUs overall, nine were present on at least fifteen sampling dates. The dinoflagellates *Heterocapsa rotundata* (OTU 3) and *Scrippsiella* sp. (OTU 5) were consistently among the most abundant OTUs. Other genera that were both abundant and consistently present included *Gyrodinium* (OTUs 13 and 16), *Tetraselmis* (OTU 19), a diatom (Mediophyceae, OTU 20), and

two prasinophytes (*Ostreococcus, OTU 6,* and *Micromonas,* OTU 12). These may comprise a 'core community' of phytoplankton species, ranging over several major taxa. Other highly abundant OTUs were present more sporadically among the different time points. Some, such as *Akashiwo sanguinea* (OTU 1) were likely bloom formers, observed abundantly on only one or several dates.

1.4.2 Choice of a cutoff for species-level identification

The clustering methods used in this paper require pre-determined sequence similarity cutoffs. The cutoff value chosen for approximate species-level distinctions had an important effect on the resulting community characterization; at lower cutoffs, sequences from multiple OTUs clustered together, 'absorbing' some of the rarer sequences that were assigned to discrete OTUs at higher cutoff values. We examined the clustering of sequences into the most abundant OTUs at different cutoffs to see whether they maintained consistent temporal patterns and found examples of three cases: some OTUs were indeed consistent regardless of cutoff (Fig. 1.3A). In a second case, a group of copepod sequences clustered together as OTU 2 at 97%, but separated into at least three OTUs with distinct abundance patterns at 99% (Fig. 1.3B). In a third case, sequences that had similar classifications and abundance patterns were assigned to different OTUs at 99%, such as OTUs 18 and 28, both classified as *Neoceratium* (Fig. 1.3C). Additional temporal sampling may provide evidence for separating these taxa. The interspecific 18S variation between copepod species (following the reproductive species concept) has been found to range from 1 to 6% (Bucklin et al, 2003). Since dynamics of three distinct copepod groups (Fig. 1.3B) were seen at 99% but not at 97%, examining temporal patterns at different cutoffs may be a useful method for determining potential species within taxa – and may be warranted especially when the focus of the study is a single taxonomic group. Within the top 25 OTUs

defined by at a similarity of 99%, five OTUs with distinct temporal patterns (three copepod and two rhizarian) appeared to be inappropriately joined into two OTUs at 97% (Figure S6).

1.4.3 Synechococcus spp. and grazer abundance at the Scripps Pier

The abundances of *Synechococcus* and their grazers at the Scripps Pier were temporally dynamic and exhibited blooms and crashes that were detectable within the weekly measurements (Fig. 1.4A). While the *Synechococcus* population generally remained below 2 x 10^5 cells ml^{-1}, their abundance exceeded this density on nine dates (2011: 7 April, 14 April, 21 April, 19 May, 7 July; 2012: 19 Jan, 10 May, 21 May, and 2 Aug). Of these, three were dates when the abundance of *Synechococcus* more than doubled within a week, and three were dates when a bloom decreased by more than half within a week. The fluctuations in abundance of *Synechococcus* did not correlate significantly with temperature (2011: Pearson's r=0.23, p-value=0.087; 2012: Pearson's r=0.27, p-value=0.05) or chlorophyll (2011: Pearson's r=0.10, p-value=0.483; 2012: Pearson's r=-0.01, p-value=0.957). *Synechococcus* grazers (as detected with intracellular prey by fluorescence microscopy) were maximally abundant on eight dates (2011: 10 May, 7 July, 21 July, 25 Aug, 15 Sept; 2012: 29 May, 21 June, and 26 July). The grazer peaks on 7 July 2011, 25 Aug 2011, and 2 Aug 2012 corresponded with peaks in *Synechococcus* abundance. Dates identified as crashes of *Synechococcus* blooms (Fig. 1.1; Fig. 1.4A) were those in which *Synechococcus* abundance declined by approximately fourfold (or greater) within the previous two weeks. *Synechococcus* abundance only had a weak significant correlation with grazer abundances (Pearson's $r = 0.26$, $p < 0.05$).

1.4.4 Identifying putative grazers of Synechococcus

Some of the increases in grazer populations that were identified by microscopy corresponded with a novel autofluorescent population observed in flow cytometric plots on

several dates, most prominently 25 Aug 2011 and 2 Aug 2012 (Fig. 1.4D,E). This population appeared near the photosynthetic picoeukaryotic algae gating region but had elevated phycoerythrin (relative orange fluorescence) signal. We hypothesized that this indicated consumption of *Synechococcus* by a photosynthetic species, and thus possible mixotrophy. To identify these populations from 2 Aug 2012, we made 18S rRNA gene clone libraries from small collections of cells that were flow-sorted from gating regions R1 and R2 (Fig. 1.4D; S3). Region R1 contained mostly dinoflagellate sequences, particularly *Gymnodinium* sp., that matched a highly abundant *Gymnodinium* OTU from the dataset (OTU 23). This OTU was present at the Scripps Pier throughout the time-series (Table 1.2), though its highest abundance was during the week *after* the *Synechococcus* bloom. Region R2 contained prasinophyte cells, mainly of the genus *Tetraselmis*, most of which matched the sequence from OTU 37 (Table 1.2). This OTU was only present on 25 Aug 2011 and 2 Aug 2012, in high abundances, indicating it may be a summer bloom-former.

A *Tetraselmis* species isolated from the Scripps Pier, which corresponded with OTU 19 (distinct from OTU 37 from region R2 above), was observed to ingest CC9311 *Synechococcus* cells when they were provided as a food source (Fig. 1.4B,C). Ingestion rates of *Synechococcus* by *Tetraselmis* increased with increasing *Synechococcus* concentration (Fig. S7). Fitted to a Michaelis-Menten equation (S4, Eqn. 2), the maximum ingestion rate (MIR) and K_{IR} (the prey concentration sustaining 1/2 MIR) of *Tetraselmis* on *Synechococcus* were 0.18 *Synechococcus* cells grazer^{-1} h^{-1} and 1.5×10^7 cells ml^{-1}, respectively (S4, Eqn. 2). The maximum clearance rate was 0.13 nl grazer^{-1} h^{-1}. A strain of *Gymnodinium* was isolated from the Scripps Pier that clustered with OTU 23 and was observed to consume *Synechococcus* cells (Yeongdu Yoo, personal communication). Given the multiple lines of evidence (cell-sorting, sequencing and

culture work) it is possible that strains of both *Gymnodinium* and *Tetraselmis* could be mixotrophic grazers of *Synechococcus* at this sampling station, presumably under very specific environmental conditions.

To further investigate putative grazers, we enriched a natural plankton community from the Scripps Pier with representatives of four *Synechococcus* clades and sequenced the community after five days. This resulted in a different shift in the community composition of microbial eukaryotes in each of the samples (compared to a control that was not enriched with *Synechococcus*). We are aware that 'surface' effects alone may cause significant shifts in the diversity of the protistan assemblage (Countway et al, 2005). The number of OTUs found within each community ranged from 158 to 237 with only 44 OTUs common to all the samples. Although our experimental treatments consisted of single enrichments for each of the *Synechococcus* clades (except for a duplicate CC9311 treatment) we observed several OTUs that experienced noticeable and consistent (including between replicates) shifts in their relative abundance compared to the un-enriched community. There was an increase in the abundance of several dinoflagellates (members of Gymnodiniphyceae and Syndiniales) and one stramenopile (Fig. 1.5). Notably, the *Gymnodinium* OTU with the greatest increase addition also matched with OTU 23, the OTU that was identified in the bulk sorts - providing another line of evidence that this species may be an important *Synechococcus* grazer at this sampling station. The relative abundance of the only abundant *Tetraselmis* OTU, which did correspond with the bloom-former OTU 37, declined in all treatments.

We further examined *Gymnodinium*-like species diversity due to its presence during *Synechococcus* blooms and because it has been demonstrated previously to exhibit mixotrophy in culture (Jeong et al, 2005, Jeong et al, 2013; Strom, 1991). Phylogenetic trees of pier sequences

that classified within Gymnodiniphycidae included a number of OTUs which clustered with the 18S sequences from two Scripps Pier *Gymnodinium*-like species isolates (Fig. 1.6). These isolates contributed to the clade Suessiaceae and may be putative grazers of *Synechococcus,* given that members of this clade appear to be abundant on the two August 2012 dates during which the R2 population appears and disappears (Fig. 1.4D). Several other Gymnodiniphycidae clades, including *Gyrodinium* sp., did not cluster with any known mixotroph sequences within the tree, though this does not rule them out as possible *Synechococcus* grazers.

1.4.5 Environmental relevance of cultured protists

We evaluated the ecological relevance of protistan strains isolated as part of other projects from the Scripps Pier by aligning the 18S rRNA gene sequences that were previously obtained from these strains (Yeongdu Yoo, Brian Palenik, Javier Paz-Yepes, unpublished data) to those from our pier time-series (Table 2). Nine strains clustered with sequences within the 100 most abundant OTUs, while two were extremely rare. The two poorly represented OTUs were both known heterotrophic nanoflagellate grazers (*Pteridomonas* and *Paraphysomonas*). In contrast, a cercozoan isolate from the Scripps Pier that grazes on *Synechococcus* (Paz-Yepes, manuscript in preparation) was more common, corresponding to OTU 262. The nine isolated strains that were common OTUs represented dinoflagellates, raphidophytes, and chlorophytes. Two species were isolated during blooms (OTU 1, *Akashiwo sanguinea,* and OTU 37, a *Tetraselmis* species) were indeed bloom formers, present only at certain sampling times but in high abundances (Table 2).

1.5 DISCUSSION

We have characterized the dynamics of the eukaryotic microbial community at the Scripps Pier to explore overall diversity and look into the factors driving *Synechococcus* ecology, which are not well understood at this site. Our sequencing-based approach uncovered eukaryotic species richness at the Scripps Pier at least tenfold greater than that described decades ago using microscopy (Allen, 1936). This corresponds with a similar theme worldwide; the 2015 TARA Oceans expedition conducted high-throughput sequencing at 47 sites (de Vargas et al, 2015) and estimated that there are approximately 150,000 eukaryotic species in the marine photic zone globally. Like the TARA data, we found that a large proportion of our sequences remained phylogenetically unclassified; the fourth most abundant OTU at the Scripps pier, present only during grazer blooms (7 July 2011, 21 July 2011, 25 Aug 2011, 21 Jun 2012), was unclassifiable beyond Eukaryota. Because microbial eukaryotes are not particularly well-represented in comparative DNA databases, it is unclear how these unclassified OTUs contribute to the coastal ecology and will require further sequencing efforts or probing with directed molecular techniques.

Despite the large range of species richness between time-points, the relative phylogenetic community structure within higher-level taxonomic designations was fairly stable, a trend that was also found between different sites in Taib et al (2013). There does appear to be a "core community" of phytoplankton OTUs (dinoflagellates, diatoms, and prasinophytes) that are present abundantly most of the time, and many of these OTUs correspond to species recorded in past work (Allen 1936, Reid 1985, Goericke 2011). However, there are many OTUs that appear to be bloom-formers: sparse or absent at most time points but highly abundant at a few. External drivers, including seasonal changes, nutrient inputs, and water transport, are likely to play a role

in the varied levels of species diversity on different dates, but the causes for most blooms remain elusive. For example, while *Akashiwo sanguinea's* dominance in the sequence data on 29 Dec 2011 was corroborated by cell counts and high chlorophyll measurements through SCCOOS, it did not correspond with any abnormal temperature, salinity, or nutrient concentration, nor did it happen during a time of year when upwelling typically occurs. Many OTUs appeared only once or twice throughout the time-series. These 'conditionally rare taxa' may have important but yet-unknown functional roles in the community (Caron & Countway, 2009; Sogin et al, 2006; Logares et al, 2015).

These results highlight the importance of temporal data in a field dominated by comparisons of spatially distinct communities sampled at a single time-point. Only three OTUs from the Scripps Pier were present throughout all time-points, but even these were not consistent in their relative temporal abundances. Thus, even if the goal is characterizing only the most abundant species, our data suggests that a single time-point is not sufficiently representative. Previous work has demonstrated rapid and large-scale changes in the composition of protistan communities during short-term experiments (Countway et al, 2005) and it is important for such shifts to be documented for an established time-series study at high temporal resolution. One instance of a dramatic shift in the eukaryotic community was observed between 2 Aug 2012 and 11 Aug 2012. The former was one of the most species-rich dates (>800 estimated OTUs) and had a *Synechococcus* bloom concurrent with an increase in grazers (observed by microscopy). By 11 Aug 2012, both *Synechococcus* and grazer abundances declined more than fivefold and the number of OTUs had decreased by an order of magnitude. These large changes in diversity and community composition within nine days highlight the need to explore time-series data at greater temporal resolution to better characterize boom-and-bust cycles. Additionally, it is presumed that

there would be a time-lag between a bloom of a given species and a detectable response in a trophically-related species, but this might only be detected with higher-frequency sampling.

The OTUs described in this study are based upon a specific method of clustering sequences (average neighbor joining with a 97% similarity cutoff). Numerous other methods have been developed to accomplish the goal of converting sequence data into ecologically meaningful information about species dynamics (Mahé et al, 2014; Edgar 2011). Previous studies have proposed sequence similarity cutoffs ranging from 95% to 99.5%, and it appears increasingly likely that the 'correct' species-level cutoff might differ between higher-level taxa (Caron et al, 2009; Massana et al, 2014; Wolf et al, 2014; Collado-Fabbri et al, 2011). We found that examining temporal patterns of OTUs at different cutoffs may be useful in evaluating how well that cutoff represents a specific population. We assumed that if two OTUs with similar taxonomic classifications also had highly similar temporal abundance patterns, the cutoff at which they clustered together might be appropriate for identifying species or ecotypes, and did find this to be the case for some OTUs at 97%. Moving forward, temporal information may aid in the definition of meaningful units of taxonomic classification by providing significantly greater coverage for any single OTU and thus justifying when taxa are split into multiple OTUs.

Another objective of this work was to closely track the *Synechococcus* population and search for factors underlying *Synechococcus* blooms and declines, with a focus on identifying grazers using both sequence data and other lines of evidence. At other locations, patterns of temperature, stratification, and nutrient concentrations are more reliably coincident with *Synechococcus* abundance (Moisan et al, 2010). However, at the Scripps Pier no significant correlations with temperature or chlorophyll (a proxy for overall phytoplankton biomass) were observed in 2011-2012, and the *Synechococcus* population did not show a strong seasonal pattern

as is observed at many other sites (Agawin et al, 1998; Shapiro and Haugen, 1988; Hunter-Cevera et al, 2016). This might be because of the relatively small temperature range at this site (13 – 23^{o}C) compared to others; additionally, *Synechococcus* has been found to only reliably correlate with temperature below 14^{o}C (Li 1998). It is also possible that our higher sampling frequency captures variation that is present but undetected at other sites.

Oceanographic features and climate events surely impact the community in many ways, but the available measurements do not explain all large community shifts. Longer-term forcing factors in the Southern California Bight include the El Nino Southern Oscillation, the Pacific Decadal Oscillation, and the North Pacific Gyre Oscillation (Di Lorenzo et al, 2008). The Scripps Pier is influenced by the equator-ward flow of the California Current and the pole-ward flow of the California Undercurrent, among others. The latter weakens in the spring and leads to seasonal variation in salinity, stratification, and nutrient availability (Lynn and Simpson, 1987). Upwelling events typically occur during the spring and summer (Dorman and Palmer, 1981), but previous work at this site has not implicated coastal upwelling in explaining chlorophyll variations (Kim et al, 2009). Much of the water movement at the Scripps Pier is attributed to wind-driven currents and this may be explanatory for some of the variations in microbial diversity. Given the lack of strong connections between *Synechococcus* abundances and oceanographic factors, we sought to explore the role of biotic factors such as grazing.

We detected a high incidence of *Gymnodinium* sequences during a *Synechococcus*-grazer bloom (in both the amplicon data and clone libraries made from the 'R2' flow cytometry sort region). The abundant *Gymnodinium* OTU sequence clustered with that of a *Gymnodinium* species known to ingest *Synechococcus* cells, and Jeong et al (2005 & 2013) have suggested that feeding might be a driving force for bloom formation in some dinoflagellates. *Gymnodinium*

appears to have a grazing threshold for *Synechococcus*; significant ingestion is only observed when prey density exceeds 20 µg C l^{-1} (Strom, 1991). Assuming 0.294 pg C *Synechococcus* $cell^{-1}$ (Strom, 1991), the *Synechococcus* concentration on 2 Aug 2012 was 67.7 µg C l^{-1}, a condition which could induce grazing.

We also found an unexpected association between *Tetraselmis* and *Synechococcus* during a bloom period. *Tetraselmis* OTU 37 was found in an unusual region of our flow-cytograms ('R1') on dates of a widely-observed 'Green Bloom,' and its 18S sequence matched that of a *Tetraselmis* isolate collected during one of these blooms. Other dates had different *Tetraselmis* OTUs present but lacked the novel population in the flow cytograms. Given that at least one *Tetraselmis* species isolated at the Scripps Pier (OTU 19) has been observed to ingest *Synechococcus* when it is presented at high density, it is possible that some *Tetraselmis* species do exhibit conditional mixotrophy at our site. We did not see an increase in *Tetraselmis* OTU sequences in an experiment where seawater was enriched with *Synechococcus* strains, despite OTU 37 being present, as might be expected if *Synechococcus* was being used as a food source. However, it is impossible to replicate the exact nutrient, temperature, and microbial community conditions that were present on the dates of the 'Green Blooms' so this does not eliminate the possibility of *Tetraselmis* mixotrophy. We recognize that the unconventional 'R1' flow cytometry population could also have been a result of physical attachment between *Tetraselmis* cells and *Synechococcus* as has been occasionally observed when they are co-cultured (Palenik lab, unpubl data), but further work is necessary to confirm or reject either hypothesis. Regardless, mixotrophy is not always considered in traditional food web schemes, but the *Gymnodinium* and *Tetraselmis* findings suggest it may play a role in consumption of *Synechococcus* at this sampling site.

Another unexpected finding was that, while the 18S sequences of many cultured dinoflagellates and green algae (isolated from the Scripps Pier) clustered with abundant OTUs, the well-characterized *Synechococcus* grazers *Pteridomonas* and *Paraphysmonas* did not, even at the relatively low sequence similarity threshold of 95%. This could indicate that efforts to isolate heterotrophic grazers select for taxa that are not dominant environmentally. It is also possible that their sequence signal was drowned out by that of organisms with higher rRNA gene copy numbers or some other DNA amplification bias (Lim et al, 1999). In reviewing various public sequence databases, Del Campo et al. (2011) observed that chrysophytes such as *Paraphysmonas*, which are observed under the microscope frequently, were not found as commonly in sequence data. This is an important consideration for future work that uses amplicon sequencing to explore these types of trophic interactions.

We have described the changing microbial community at the Scripps Pier over two years and found it to be diverse and highly dynamic, with only a few species consistently present. Amplicon sequencing captured blooms of certain microbial eukaryotes. Our results clearly emphasize the importance of temporal studies in describing microbial communities.

1.6 FIGURES AND TABLES

Table 1.1. *Synechococcus* and *Synechococcus* grazer density for the sampling dates for which 18S sequencing was conducted. *Synechococcus* cells were counted using flow cytometry and their grazers were counted using microscopy (see methods). Temperature and chlorophyll data were obtained from the SCCOOS automated shore station database (http://www.sccoos.org/data/autoss/).

Date	*Synechococcus* (cells/ml)	*Syn.* grazers (cells/ml)	Temperature (C)	Chlorophyll (μg/L)
14 Apr 2011	288283	255	15.39	1.02
5 May 2011	137443	766	17.27	1.92
7 Jul 2011	281097	8303	16.7	1.6
21 Jul 2011	53594	8494	15.05	0.97
25 Aug 2011	134166	12646	16.4	0.65
15 Sept 2011	36227	1597	18.53	0.64
17 Nov 2011	29254	1980	15.53	0.52
29 Dec 2011	72740	383	14.02	1.78
26 Jan 2012	58980	64	14.06	1.1
1 Mar 2012	28349	1022	14.11	1.02
10 May 2012	247034	383	18.18	0.59
29 May 2012	128537	1469	19.22	0.54
14 Jun 2012	61794	128	17.57	1.37
21 Jun 2012	55069	1661	19.4	1.48
2 Aug 2012	230162	13667	21.49	1.7
11 Aug 2012	26528	703	21.69	1.72
25 Oct 2012	43957	1213	18.26	2.01

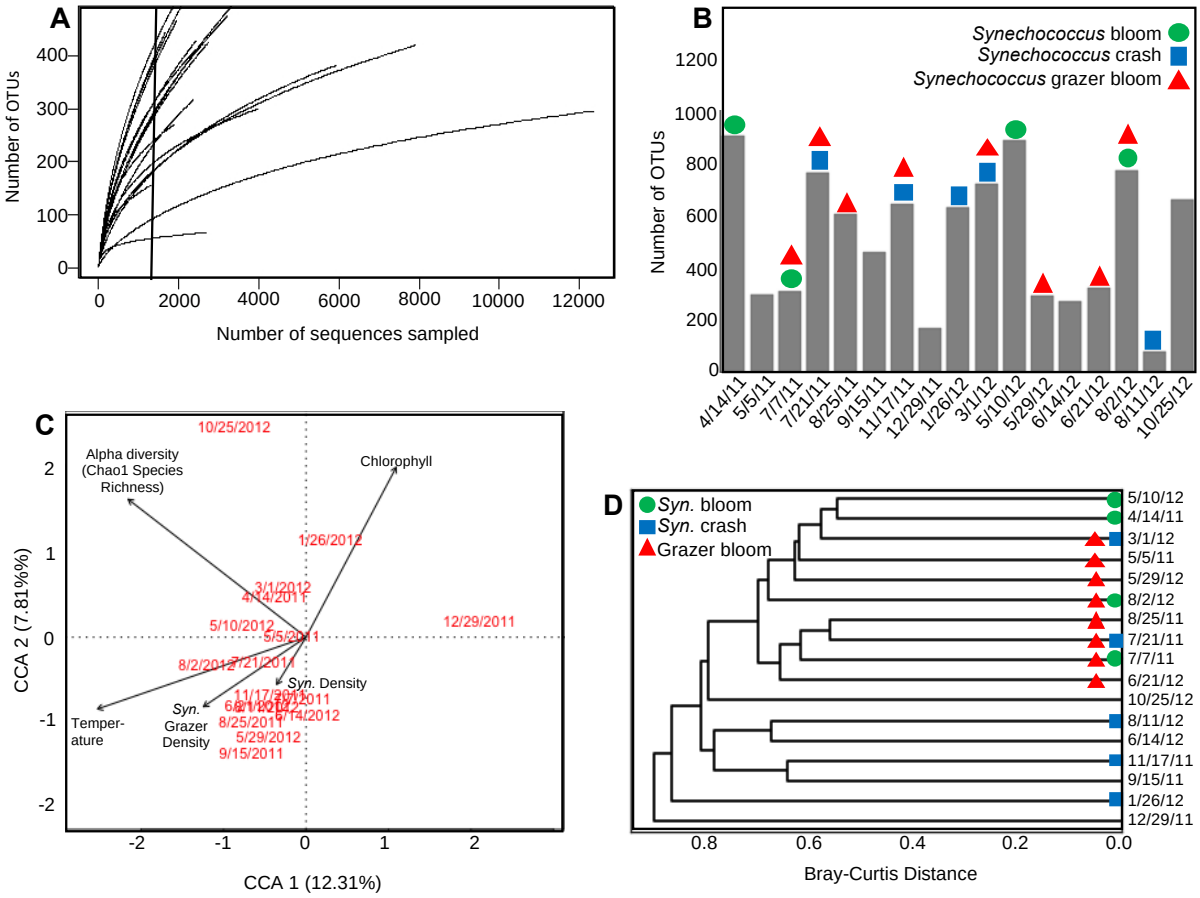


Figure 1.1. (A) Rarefaction curves for each of the samples (with OTUs assigned at a 97% similarity cutoff). Vertical line represents level of sub-sampling (1308 sequences) that was used for subsequent analysis. (B) Estimated number of OTUs, based on the Chao1 richness indicator, present at each time point. (C) Canonical Correspondence Analysis of communities and selected environmental variables. (D) Dendrogram clustering communities at each sampling point based on unweighted Bray-Curtis dissimilarity.

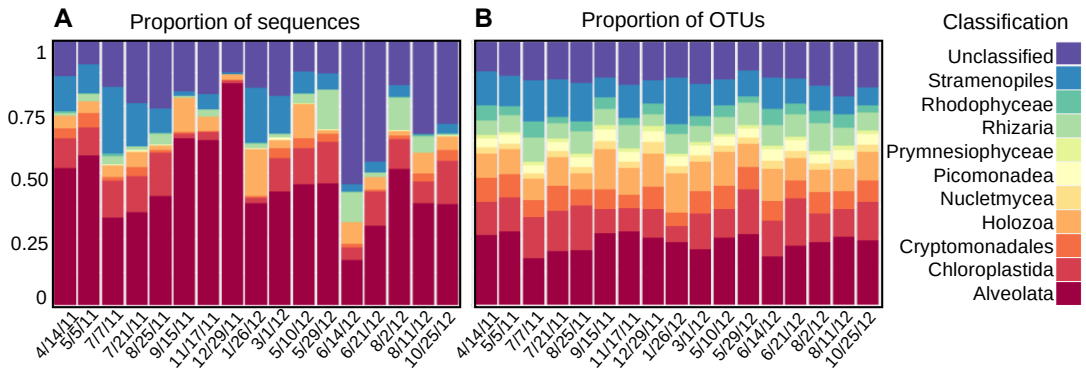


Figure 1.2. Temporal eukaryotic diversity at the SIO pier. (A) Proportion of total sequences falling within each of the taxa, representing relative abundance of those taxa. (B) Proportion of total OTUs falling within each of the taxa, representing relative species richness within taxa.

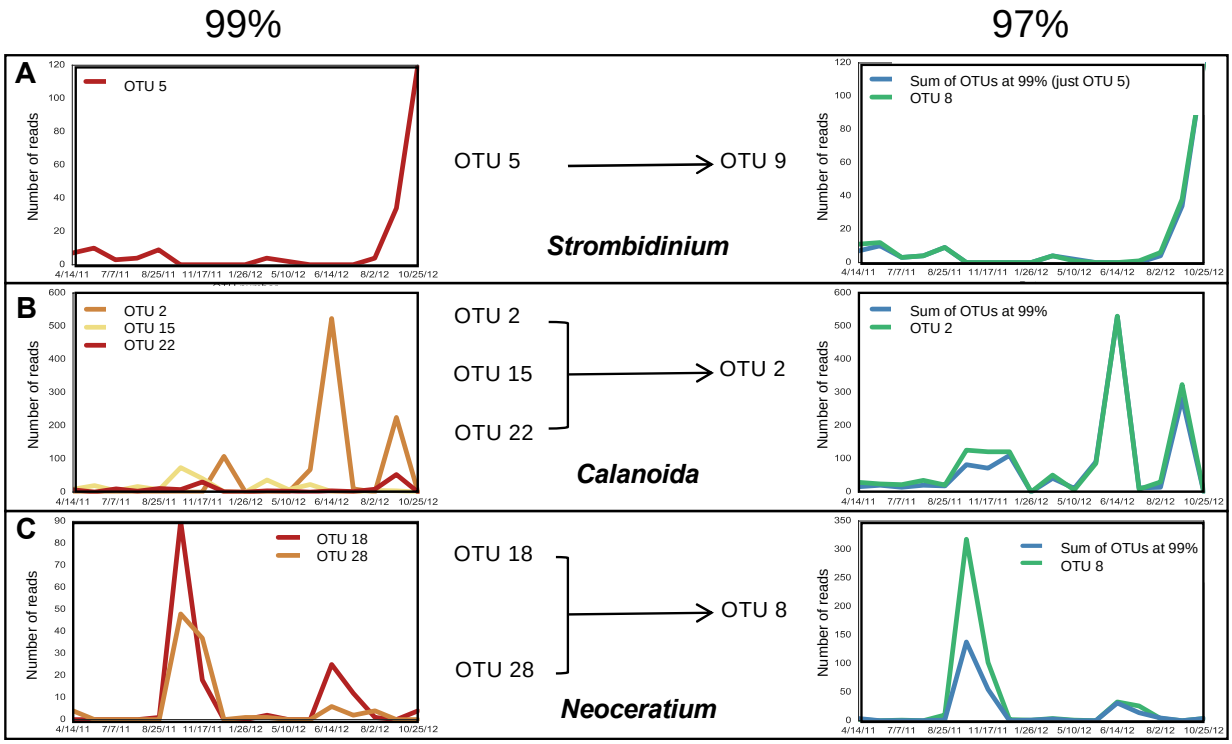


Figure 1.3. Examples of OTU abundances throughout the time series, with OTUs at 99% (left) and 97% (right) that presumably represent the same organism. Leftmost graphs show an OTU or multiple OTUs at 99%, while rightmost graphs overlay the sum of the OTU abundances from the left graph (blue) with a corresponding OTU at 97% (green). (A) A case where an OTU remains consistent at both cutoffs. (B) A case where three separate OTUs with distinct temporal patterns are assigned at 99%, but the sum of their abundances is roughly equal to the abundance pattern of a similarly classified OTU at 97%. In this case we assume the OTUs assigned at 99% represent different species due to their very different abundance patterns, and use of a 97% cutoff masks this diversity. (C) A case where two separate OTUs with very similar temporal patterns are assigned separately at 99% but the sum of their abundances closely matches a single OTU as assigned at 97%. In this case it appears that use of a 99% cutoff creates "false diversity." The discrepancy in abundance between OTU 8 and the sum of OTUs 18 and 28 is due to many other sequences being "absorbed" into OTU 8 when clustered at 97%.

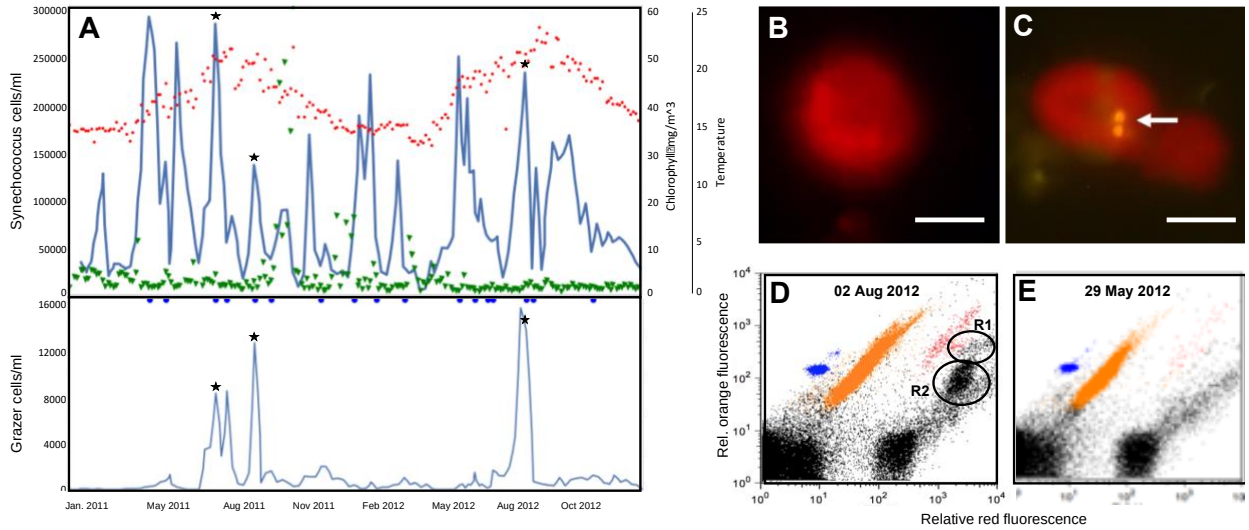
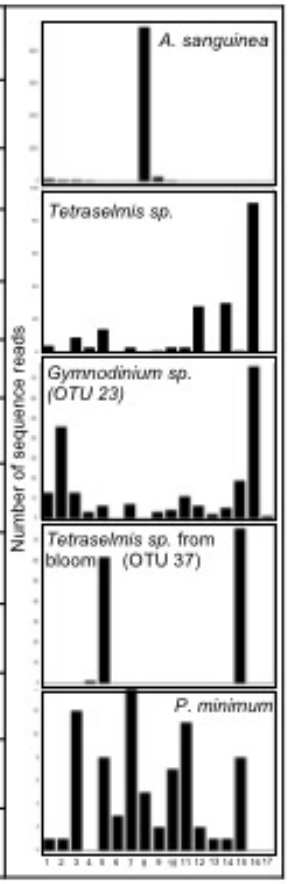


Figure 1.4. (A) *Synechococcus* and *Synechococcus* grazer density in cells/ml in 2011 and 2012. *Synechococcus* cells were counted using flow cytometry; grazers were counted using epifluorescence microscopy as described in methods. Marked time points (blue circles) show dates where 18S environmental sequences were analyzed. Chlorophyll (green triangles) and temperature (red dots) are shown along with *Synechococcus* abundances. *Synechococcus* abundance exceeded 2 x 10^5 cells ml^{-1} on nine dates (2011: 7 April, 14 April, 21 April, 19 May, 7 July; 2012: 19 Jan, 10 May, 21 May, and 2 Aug). Stars denote dates that have both *Synechococcus* and grazer blooms. (B) Epifluorescence microscopy image of a single *Tetraselmis* cell isolated from the SIO pier. Scale bar = 5 □m. (C) Image of a *Tetraselmis* cell that has been fed *Synechococcus*. Arrow points to *Synechococcus* cells that have been ingested by the *Tetraselmis* cells. (D,E) Examples of flow cytometry plots from a date with novel cell populations (R1 and R2, circled), 2 Aug 2012 (D), compared to a more 'typical' date, 29 May 2012 (E). The blue region marks the control beads, red region denotes putative cryptophytes, and the orange region represents the *Synechococcus* population.

Table 1.2. Names of selected Scripps Pier isolates, the OTU from the environmental 18S sequences that the isolate sequence clustered with at 97% (column 2), and sequence abundances throughout time for five of these OTUs (bar graphs). Dates 1-17 correspond with dates from other figures: in order, 14 Apr 2011; 5 May 2011; 7 Jul 2011; 21 Jul 2011; 25 Aug 2011; 15 Sept 2011; 17 Nov 2011; 29 Dec 2011; 26 Jan 2012; 1 Mar 2012; 10 May 2012; 29 May 2012; 14 Jun 2012; 21 Jun 2012; 2 Aug 2012; 11 Aug 2012; 25 Oct 2012.

Species in culture	**OTU abundance (throughout time series)**
Akashiwo sanguinea	999
Heterocapsa rotundata	1413
Scrippsiella trochoidea	840
Ceratium sp.	508
Tetraselmis sp. (1)	190
Gymnodinium sp.	215
Tetraselmis sp. from August 2012 bloom	139
Prorocentrum minimum	76
Prorocentrum micans	47
Pteridomonas sp.	1
Paraphysmonas sp.	1
Cercozoan species	9



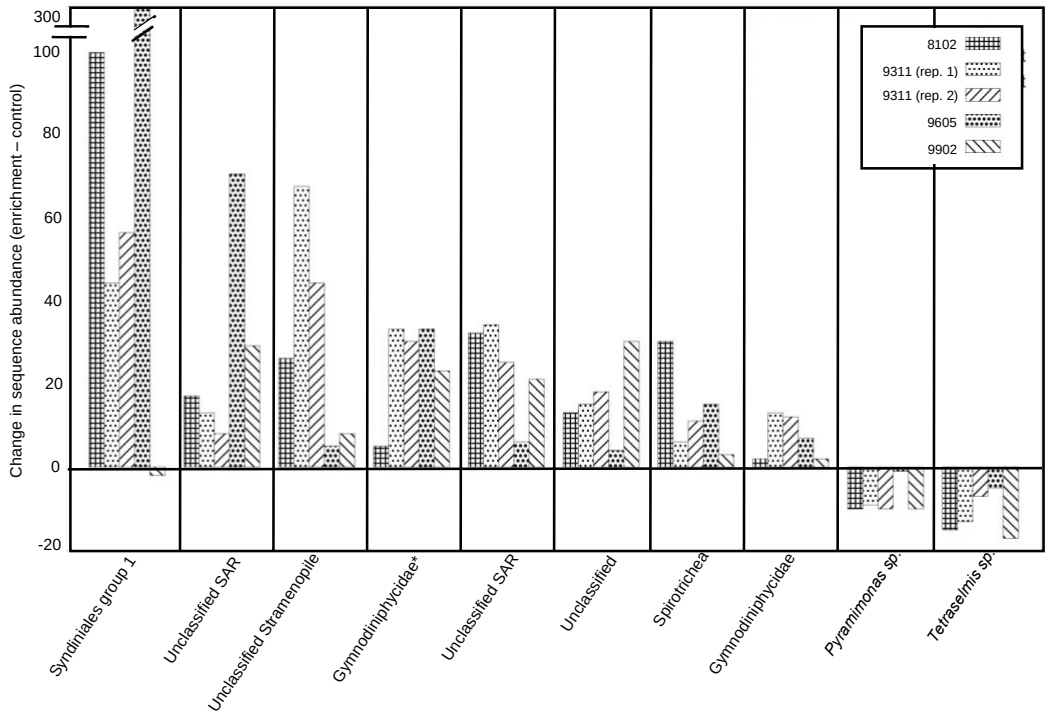


Figure 1.5 Changes in sequence abundance of selected OTUs in seawater enriched with four different *Synechococcus* species with respect to unenriched control. OTUs shown had the largest overall increases or decreases in abundance compared to the control. Starred Gymnodiniphycidae OTU corresponds with the *Gymnodinium* OTU found in the R2 flow cytometry region from 2 Aug 2012. SAR = Superkingdom of Stramenopiles, Alveolates, Rhizarians. Patterns represent the *Synechococcus* strain added to the enrichment.

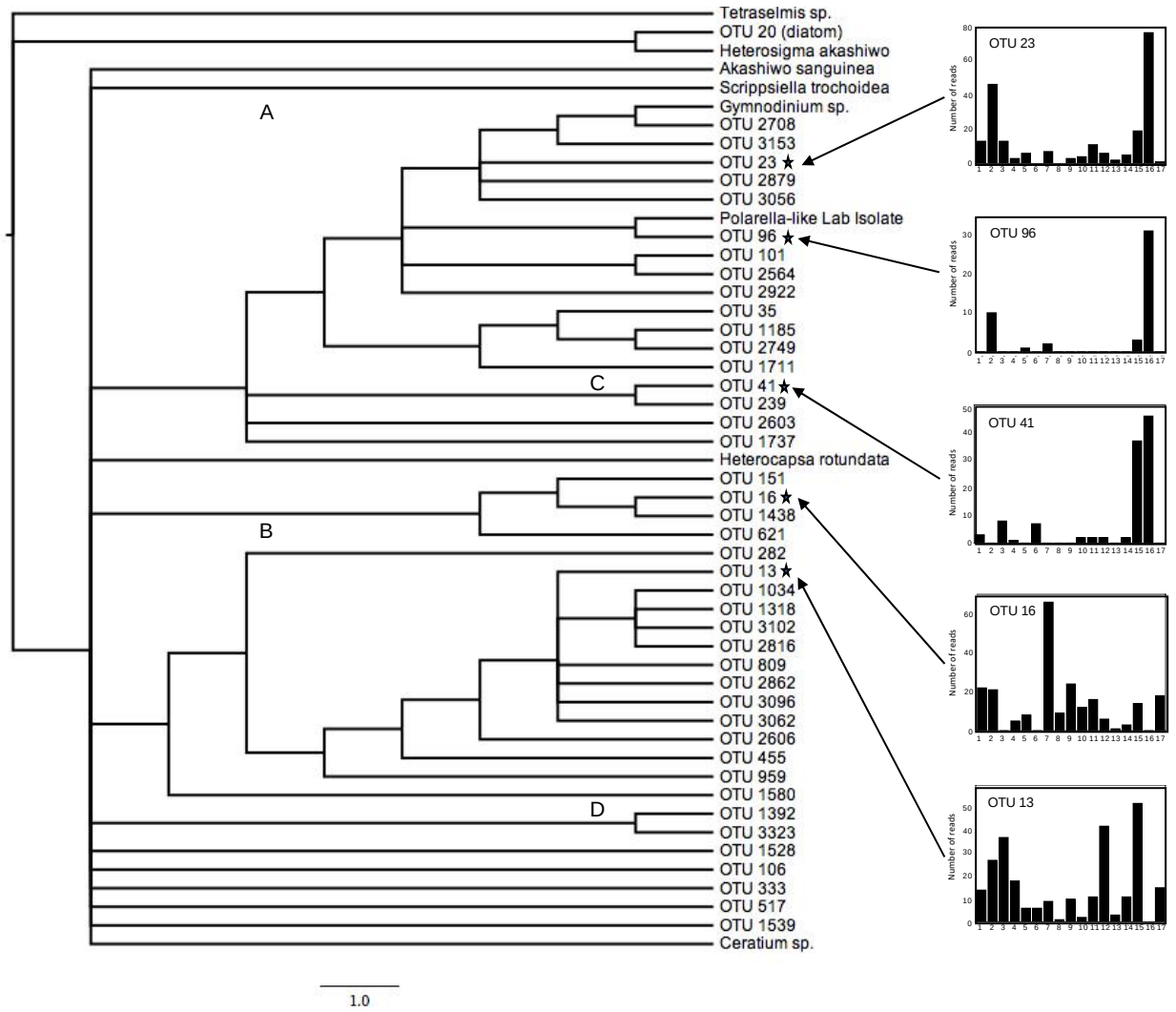


Figure 1.6 Consensus maximum-likelihood tree of Gymnodiniphycidae sequences found at pier as well as sequences of dinoflagellate lab isolates. Colors represent four taxonomic groups as described in the SILVA database, *Suessiaceae* (pink), *Gyrodinium* (yellow), and two other separately designated clades (green and blue). Arrows point to most abundant OTUs (starred) and show their abundance patterns over time. Dates 1-17 correspond with dates from other figures: in order, 14 Apr 2011; 5 May 2011; 7 Jul 2011; 21 Jul 2011; 25 Aug 2011; 15 Sept 2011; 17 Nov 2011; 29 Dec 2011; 26 Jan 2012; 1 Mar 2012; 10 May 2012; 29 May 2012; 14 Jun 2012; 21 Jun 2012; 2 Aug 2012; 11 Aug 2012; 25 Oct 2012.

1.7 ACKNOWLEDGEMENTS

Funding for this project was provided through a National Science Foundation grant (DEB-1233085) to BP, as well as National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIP) (NRF-2015M1A5A1041808) award to YDYOO. We thank M Carter and M Hilbern for collecting access to the Scripps Pier Chlorophyll Program dataset, funded by NOAA and COTS through the Southern California Coastal Ocean Observing System (SCCOOS). We thank Mike Landry (SIO) for use of his automated microscopy system used for obtaining images for *Synechococcus* grazer counts.

Chapter 1, in full, is a reprint of the material as it appears in ISME 2018. Nagarkar, Maitreyi; Countway, Peter; Yoo, Yeong Du; Daniels, Emy; Poulton, Nicole; Palenik, Brian. 2016. The dissertation author was the primary author on this paper.

1.8 REFERENCES

Agawin NSR, Duarte CM, Agustí S. (1998). Growth and abundance of Synechococcus sp. in a Mediterranean Bay: seasonality and relationship with temperature. Mar Ecol-Prog Ser 170:45-53.

Agawin NSR, Duarte CM, Agustí S. (2000). Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. Limnol Oceanogr 45(3): 591–600.

Allen WE. (1936). Occurrence of marine plankton diatoms in a ten-year series of daily catches in southern California. Am J Bot 60-63.

Apple JK, Strom SL, Palenik B, Brahamsha B. (2011). Variability in protist grazing and growth on different marine Synechococcus isolates. Appl Environ Microbiol 77(9):3074–3084.

Bucklin A, Frost B, Bradford-Grieve J, Allen L, Copley N. (2003). Molecular systematic and phylogenetic assessment of 34 calanoid copepod species of the Calanidae and Clausocalanidae. Marine Biol 142(2):333-343.

Caron DA, Countway PD. (2009). Hypotheses on the role of the protistan rare biosphere in a changing world. Aquat Microb Ecol 57:227-238

Caron DA, Countway PD, Savai P, Gast RJ, Schnetzer A, Moorthi SD *et al*. (2009). Defining DNA-based operational taxonomic units for microbial-eukaryote ecology. Appl Environ Microbiol 75(18):5797–5808.

Cheung MK, Au CH, Chu KH, Kwan HS, Wong CK. (2010). Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. ISME J 4(8):1053-1059.

Christaki U, Courties C, Karayanni H, Giannakourou A, Maravelias C, Kormas KA *et al*. (2002). Dynamic characteristics of Prochlorococcus and Synechococcus consumption by bacterivorous nanoflagellates. Microb Ecol 43(3):341–352.

Cleary AC, Durbin EG. (2016). Unexpected prevalence of parasite 18S rDNA sequences in winter among Antarctic marine protists. J Plankton Res fbw005.

Collado-Fabbri S, Vaulot D, Ulloa O. (2011). Structure and seasonal dynamics of the eukaryotic picophytoplankton community in a wind-driven coastal upwelling ecosystem. Limnol Oceanogr 56(6):2334–2346.

Collier JL, Palenik B. (2003). Phycoerythrin-containing picoplankton in the Southern California Bight. Deep-Sea Res PT II 50(14-16):2405–2422.

Countway PD, Gast RJ, Savai P, Caron DA. (2005). Protistan diversity estimates based on 18S rDNA from seawater incubations in the western North Atlantic. J Eukaryot Microbiol 52:95–106.

Countway PD, Vigil PD, Schnetzer A, Moorthi SD, Caron DA (2010) Seasonal analysis of protistan community structure and diversity at the USC microbial observatory (San Pedro Channel, North Pacific Ocean). Limnol Oceanogr 55:2381-2396.

Del Campo J, Massana R. (2011). Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. Protist 162(3):435-448.

De Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares, R *et al*. (2015). Eukaryotic plankton diversity in the sunlit ocean. Science 348(6237):1261605.

Di Lorenzo E, Schneider N, Cobb KM, Franks PJS, Chhak K, Miller AJ *et al.* (2008). North Pacific Gyre Oscillation links ocean climate and ecosystem change. Geophysical Research Letters, 35(8).

Dorman CE & Palmer DP. (1981). Southern California summer coastal upwelling. Coastal upwelling, 44-56.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27(16):2194-2200.

Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG and Sogin ML. (2013). Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol, 4: 1111–1119.

Fenchel T. (1982). Ecology of heterotrophic microflagellates. IV. Quantitative occurrence and importance as bacterial consumers. Mar Ecol Prog Ser 9(1):35-42.

Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N *et al*. (2013). Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. PNAS 110(24):9824–9829.

Goericke R. (2011). The structure of marine phytoplankton communities- patterns, rules and mechanisms. Cal Coop Ocean Fish 52:182–197.

Guillard RR. (1975). Culture of phytoplankton for feeding marine invertebrates. Culture of marine invertebrate animals. Springer US, pp 29-60.

Hunter-Cevera KR, Neubert MG, Olson RJ, Solow AR, Shalapyonok A, Sosik HM. (2016). Physiological and ecological drivers of early spring blooms of a coastal phytoplankter. Science, 354(6310):326-329.

Jeong HJ, Park JY, Nho JH, Park MO, Ha JH, Seong KA *et al.* (2005). Feeding by red-tide dinoflagellates on the cyanobacterium *Synechococcus*. Aquat Microb Ecol 41(2):131-143.

Jeong HJ, Du Yoo Y, Lee KH, Kim TH, Seong KA, Kang NS *et al.* (2013). Red tides in Masan Bay, Korea in 2004–2005: I. Daily variations in the abundance of red-tide organisms and environmental factors. Harmful Algae 30:S75-S88.

Jones E, Oliphant E, Peterson P, et al. (2001). SciPy: Open Source Scientific Tools for Python. http://www.scipy.org/ [Online; accessed 2017-08-08].

Kim HJ, Miller AJ, McGowan J, Carter ML. (2009). Coastal phytoplankton blooms in the Southern California Bight. Prog Oceanogr 82(2), 137-147.

Lim EL, Dennett MR, Caron DA. (1999). The ecology of *Paraphysomonas imperforata* based on studies employing oligonucleotide probe identification in coastal water samples and enrichment cultures. Limnol Oceanogr 44:37-51.

Logares R, Mangot JF, Massana R. (2015). Rarity in aquatic microbes: placing protists on the map. Res Microbiol 166(10):831-841.

Lynn RJ & Simpson JJ. (1987). The California Current System: The seasonal variability of its physical characteristics. Journal of Geophysical Research: Oceans, 92(C12):12947-12966.

Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M. (2014). Swarm: robust and fast clustering method for amplicon-based studies. PeerJ 2:e593.

Massana R, Balagué V, Guillou L, Pedros-Alio C. (2004). Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. FEMS Microbiol Ecol 50(3):gp231-243.

Massana R, del Campo J, Sieracki ME, Audic S, Logares R. (2014). Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. ISME J 8(4):854–66.

Moisan TA, Blattner KL, Makinen CP. (2010). Influences of temperature and nutrients on Synechococcus abundance and biomass in the southern Mid-Atlantic Bight. Cont Shelf Res 30(12):1275–1282.

Moon-van der Staay SY, De Wachter R, Vaulot D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409(6820):607–610.

Not F, Latasa M, Scharek R, Viprey M, Karleskind P, Balagué V *et al.* (2008). Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. Deep-Sea Res Pt I 55(11):1456-1473.

Oksanen J, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R et al. (2013). vegan: Community Ecology Package. R package version 2.0-10. R Package version 1 (http://cran.r-project.org).

Olson RJ, Chisholm SW, Zettler ER, Armbrust E. (1990). Pigments, size, and distributions of Synechococcus in the North Atlantic and Pacific Oceans. Limnol Oceanogr 35(1):45-58.

Palenik BP. (2000). Picophytoplankton seasonal cycle at the SIO pier, La Jolla, California. J Phycol 36(s3):53-53.

Partensky F, Blanchot J, Vaulot D. (1999). Differential distribution and ecology of Prochlorococcus and Synechococcus in oceanic waters: a review. Bull Inst Oceanogr 19(19):457–475.

Pomeroy LR. (1974). The ocean's food web, a changing paradigm. Bioscience 24(9):499-504.

Potter D, Lajeunesse T, Saunders G, Anderson R. (1997). Convergent evolution masks extensive biodiversity among marine coccoid picoplankton. Biodivers Conserv 107:99–108.

Prokopowich CD, Gregory TR, Crease TJ. (2003). The correlation between rDNA copy number and genome size in eukaryotes. Genome 46(1), 48-50.

Pruesse E, Peplies J, Glöckner FO. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28(14):1823-1829.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al.* (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids res gks1219.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reid FM, Lange CB, White MM. (1985). Microplankton species assemblages at the Scripps Pier from March to November 1983 during the 1982-1984 El Nino event. Botanica Marina, 28(10): 443-452.

Rinke C, Lee J, Nath N, Goudeau D, Thompson B, Poulton N *et al.* (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS–based single-cell genomics. Nature protocols 9(5):1038-1048.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75(23):7537-7541.

Schloss PD, Gevers D, Westcott SL. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PloS ONE 6:e27310.

Seenivasan R, Sausen N, Medlin LK, Melkonian M. (2013). Picomonas judraskeda Gen. Et Sp. Nov.: The First Identified Member of the Picozoa Phylum Nov., a Widespread Group of Picoeukaryotes, Formerly Known as "Picobiliphytes." PLoS ONE 8(3).

Shapiro LP, Haugen EM. (1988). Seasonal distribution and temperature tolerance of Synechococcus in Boothbay Harbor, Maine. Estuarine, Coastal and Shelf Science, 26(5): 517-525.

Sherr BF, Sherr EB, Fallon RD. (1987). Use of monodispersed, fluorescently labeled bacteria to estimate in situ protozoan bacterivory. Appl Environ Microbiol 53(5):958-965.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". PNAS 103(32):12115-12120.

Stoeck T, Bass D, Nebel M, Christen R, Jones MD, Breiner HW *et al.* (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol 19(s1):21-31.

Strom SL. (1991). Growth and grazing rates of the herbivorous dinoflagellate Gymnodinium sp. from the open subarctic Pacific Ocean. MEPS 103-113.

Tai V, Palenik B. (2009). Temporal variation of Synechococcus clades at a coastal Pacific Ocean monitoring site. ISME J 3(8):903-915.

Taib N, Mangot JF, Domaizon I, Bronner G, Debroas D. (2013). Phylogenetic affiliation of SSU rRNA genes generated by massively parallel sequencing: new insights into the freshwater protist diversity. PLoS One 8(3):e58950.

Taylor AG, Landry MR, Selph KE, Wokuluk JJ. (2014). Temporal and spatial patterns of microbial community biomass and composition in the Southern California Current Ecosystem. Deep-Sea Res PT II 112:117–128.

Waterbury JB, Watson SW, Guillard RR, Brand LE. (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. Nature 277(5694):293-294.

Wolf C, Frickenhaus S, Kilias ES, Peeken I, Metfies K. (2014). Protist community composition in the Pacific sector of the Southern Ocean during austral summer 2010. Polar Biol 37(3):375–389.

Worden AZ, Nolan JK, Palenik B. (2004). Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. Limnol Oceanogr 49(1):168-179.

Zwirglmaier K, Spence E, Zubkov MV, Scanlan DJ, Mann NH. (2009). Differential grazing of two heterotrophic nanoflagellates on marine Synechococcus strains. Environmen Microbiol 11(7):1767–1776.

CHAPTER 2:

Spatial and temporal variations in *Synechococcus* microdiversity in the Southern California Coastal ecosystem

2.1 ABSTRACT

The *Synechococcus* population at the Scripps Institution of Oceanography (SIO) pier in La Jolla, CA has been monitored at the genus level on a weekly basis for over a decade. Large increases in abundance, termed "blooms", are typically observed in the spring and summer and decline within weeks. Here we used amplicon sequencing of the ITS (internal transcribed spacer) region to examine the microdiversity of this cyanobacterial genus during blooms at the SIO pier as well as further offshore in the Southern California coastal ecosystem (CCE). These analyses revealed numerous amplicon sequence variants (ASVs) to be present. We found that the *Synechococcus* composition can change over the course of blooms. The ratio between the typically dominant clades at our site (I:IV) was different during each bloom and shifted over the course of some, but not all, blooms. We also found that in 2016 there was a highly anomalous bloom, both in its overall *Synechococcus* abundance and in terms of the *Synechococcus* clades present. Members of the generally oligotrophic clade II that were not detected during any of the other blooms comprised an unexpectedly large proportion of this bloom. The dominant ASVs found further offshore and in the California Current were the same as those at the pier, but we did observe members of some more oligotrophic clades along with depth variation in the *Synechococcus* diversity. We also observed that the dominant sequence variant switched during the peak of multiple *Synechococcus* blooms, with this switch occurring in multiple clades. We were able to repress the presence of one variant by nitrogen-limiting a *Synechococcus* culture, while control cultures contained both variants – thus we believe this apparent ASV switch to be a physiological response rather than a change in the dominant population. Our results indicate that there are hidden dynamics of *Synechococcus* microdiversity with distinct spatial and temporal patterns.

2.2 INTRODUCTION

Synechococcus is a globally-important genus of cyanobacteria that is responsible for a significant amount of primary production, especially at coastal sites (Agawin et al., 2000; Collier et al., 2003). Originally discovered using because of its unique autofluorescence signature, (Waterbury et al., 1979), *Synechococcus* was soon recognized to be highly ubiquitous throughout the ocean using shipboard flow cytometry (Olson et al., 1985) and to consist of multiple clades with differing environmental niches and capabilities (Ferris & Palenik, 1998). These different species- or strain-level populations offer insights into the adaptive strategies of marine bacteria to the wide range of changing environments they encounter in the ocean.

Currently the genus *Synechococcus* is thought to comprise at least 18 well-defined clades which vary in genome size/composition, nutrient requirements, and geographic distribution (Ahlgren & Rocap, 2012; Choi et al., 2014; Rocap et al., 2002; Sohm et al., 2016; Zwirglmaier et al., 2007). Sequence-based phylogenies have been developed using multiple genetic markers including the 16S ribosomal subunit (Rocap et al., 2002) which is commonly used for the classification of bacteria. Other commonly used markers such as the RNA polymerase gene (*rpoC1*) (Palenik, 1994; Palenik & Haselkorn, 1992; Toledo et al., 1999), photosystem I gene PsbA (Zeidner et al., 2003), and the nitrate reductase gene narB (Paerl et al., 2011) have been developed to provide greater taxonomic resolution. Here, we utilize a region of the internal transcribed spacer (ITS) that has been demonstrated to provide high resolution in differentiating between subtypes within the *Synechococcus* subcluster 5.1 and is also short enough to use with high-throughput sequencing (Choi et al., 2014). Sequencing of this ITS region has shown some

of the previously well-supported clades to be robust and revealed several novel clades (Choi & Noh, 2009).

The temporal and spatial patterns of *Synechococcus* abundance have been characterized in a number of studies. At some sites there is a seasonal pattern of high abundance in the summer months (Agawin et al., 1998; Hunter-Cevera et al., 2016; Patterson, 1998; Xia et al., 2015), whereas at other sites (including our own) the *Synechococcus* population exhibits short-lived blooms in the spring or summer (Robidart et al., 2011; Tai & Palenik, 2009). In either case, relationships between *Synechococcus* abundance and temperature (Agawin et al., 2000; Robidart et al., 2011), nitrate (Collado-Fabbri et al., 2011; Glover et al., 1988; Rajaneesh & Mitbavkar, 2013), and upwelling (Collier et al., 2003) have been identified. These relationships have been found spatially as well; molecular studies have demonstrated that different clades of *Synechococcus* occupy different geographic and depth ranges. Sohm et al. (2016) found that different *Synechococcus* clades dominated different oceanic niches and defined four regimes based on temperature, macronutrients, and iron availability. It is well known that Clades I and IV tend to be dominant in coastal environments, including the coastal California ecosystem (Zwirglmaier et al., 2007) and the Scripps pier specifically (Tai et al., 2011). Many of the other marine *Synechococcus* clades, including clades II and III, have been observed more abundantly in oligotrophic, open-ocean environments (Sohm et al., 2016; Zwirglmaier et al., 2007). Certain clades also demonstrate monophyletic environmental adaptations such as motility (Toledo et al., 1999). The question of why certain, evolutionarily distinct clades consistently co-occur (such as clades I and IV at our site) remains, and more detailed knowledge of their growth patterns along with how they are affected by abiotic conditions will likely deepen our understanding.

Recent sequence analysis methods have indicated that microdiversity within the broader clades could be significant and that in some cases there are spatial or temporal patterns among very closely related members. This microdiversity, defined in a recent review as "microbial sub-taxa within larger phylogenetic groups that have distinct niche space and greater than 97% 16S rRNA gene similarity" (Larkin & Martiny, 2017), is identifiable using amplicon sequencing of the ITS region; Choi et al (2014) found several hundred distinct representative sequences of *Synechococcus* subcluster 5.1 from oligotrophic sites in the northwestern Pacific. Mackey et al (2017) observed a consistent seasonal succession of several different *Synechococcus* oligotypes over the course of the summer at several ocean and estuarine sites in Little Sippewissett, Cape Cod, MA. This pattern included changes in the dominant clades present, but also a switch in the dominant oligotype within some of the clades.

It is not always known whether microbial strains with high sequence identity in their rRNA sequences represent actual functional diversity, though there is evidence of fine-scale patterning of phylogenetically distinct populations (Acinas et al., 2003; Lee et al., 2019). Environmental sequencing of *Prochlorococcus* shows clusters of ecotypes that have >98% similarity in the ITS region, and single-celled whole-genome sequencing of some of these indicates that these subpopulations have distinct "genomic backbones" (Kashtan et al., 2014). Even subpopulations within the same clade contained unique gene cassettes, which may confer differential adaptation in functions like redox stress response and outer membrane modification (Kashtan et al., 2014). Members of the same *Synechococcus* clade have also been shown to have varying genome size and GC content as well as distinct geographic distributions (Lee et al., 2019). Similarly, at the SIO pier, detection of horizontally-acquired genes that conferred copper and oxidative stress tolerance in *Synechococcus* clade I varied with time, indicating variation in

the functional abilities within members of a single clade (Stuart et al., 2013). Much remains to be explored on the role of physiochemical forces on phytoplankton microdiversity, including the spatial scales on which patterning can be detected. Recent evidence from the Tara Oceans Expedition (Farrant et al., 2016) and other studies (Paerl et al., 2011; Slack et al., 2014) suggest that shifts in microdiversity over short geographic distances exist in the marine environment.

In this paper, we apply high-throughput sequencing of the 16S-23S Internal Transcribed Spacer (ITS) region (Choi, Hoon Noh, & Lee, 2014) to examine of *Synechococcus* microdiversity in the southern California bight both temporally (over the course of *Synechococcus* blooms at a single site, the SIO Pier) and spatially (comparing four different sampling sites in the California Current ecosystem). Based on past knowledge of differential clade abundances over time at our site (Tai et al., 2011; Tai & Palenik, 2009), we expect a change in the clade I:IV ratio between the onset and termination of a *Synechococcus* bloom, with clade I more abundant prior to the bloom. We further hypothesize that there might be temporal niches for the co-occurring *Synechococcus* ASVs even within the dominant clades. Spatial differentiation is also likely to occur and based on previous studies we would expect clades I and IV to be the most abundant at the Scripps pier, but other clades (including II and III) to be more abundant further offshore.

2.3 METHODS

2.3.1 Sample collection

We collected a surface seawater sample by lowering a bucket at the end of the Scripps Institution of Oceanography pier (32º87'N, 117º26'W) weekly (before 2014) or bi-weekly (after 2014). Seawater temperature from selected sampling points is provided in S1. 500 ml of

seawater was filtered in triplicate onto 47-mm diameter 0.2-µm Supor (Pall corporation, Port Washington, NY, USA) membrane filters for total DNA extraction, and filters were stored at -80°C until further use. For flow cytometry counts, we preserved 1 ml seawater in 0.25% glutaraldehyde (Sigma-Aldrich, St Louis, MO, USA), incubated at room temperature for 10 min, and then stored cryovials at -80°C until further use.

Samples from the CCE LTER P1604 cruise were obtained through CTD casts at different experimental cycles and depths (specified in S2) and processed in two different ways. Some samples were filtered shipboard as described above and others were processed slightly differently (Valencia et al, unpubl data): 280 ml (200-µm Nitex screen) or 650 ml of seawater (500-µm Nitex screen) were prescreened to remove mesozooplankton prior to filtration through a 25-mm diameter 0.2-µm Supor membrane filters (Pall Corporation, Port Washington, NY, USA). Once the water was filtered, the filters were folded in half, placed in 2-ml screw-cap cryogenic vials, flash-frozen in liquid nitrogen, and stored at -80°C until analysis.

2.3.2 Synechococcus counts

We used flow cytometry to measure *Synechococcus* abundance at each sampling point. For pier samples, the glutaraldehyde-fixed samples were thawed in the dark and then mixed with 18 µl of 0.94 µm green fluorescent beads (Duke Scientific Corporation, Palo Alto, CA, USA). These were then run on a BD FACsort (Beckton Dickinson, Franklin Lakes, New Jersey, USA) for approximately five min at the HI setting. We gated a specific region on the flow cytograms to represent *Synechococcus* cells and calculated *Synechococcus* density by normalizing the event counts to the volume of sample run and the counts of beads. Samples were typically processed 1-6 months after they were collected. *Synechococcus* counts from sampling dates used for amplicon sequencing are shown in S1.

Samples from the CCE LTER P1604 cruise were treated slightly differently. They were preserved in paraformaldehyde and processed as described in Selph et al. (2011).

2.3.3. DNA extraction

For all pier samples and selected cruise samples (as specified in S2), frozen filter samples were cut into small pieces using a clean razor and the pieces were divided among two 2 ml microcentrifuge tubes. 560 µl of TE (50 mM Tris, 20 mM EDTA) and 80 µl 100 mg/ml lysozyme were added to each tube and tubes were incubated for 30 min at 37ºC. Subsequently 80 µl of 10% sodium dodecyl sulfate and 80 µl of 10 mg/ml Proteinase K were added and tubes were incubated for 2.5 h at 55ºC. Then 16 µl of 10 mg/ml RNAse A was added and tubes were incubated an additional 30 min at 37ºC. An equal volume of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) was added to each sample and the aqueous layer was pipetted into a new tube; this was repeated once and then once with Chloroform:Isoamyl Alcohol (24:1). Finally, DNA was eluted using the Qiagen DNEasy Blood and Tissue kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions (with reagent volumes scaled up when appropriate). DNA was stored at -20ºC until it was sent out for amplicon sequencing.

Environmental DNA from the remaining cruise samples was extracted using the NucleoMag 96 Plant kit (Macherey Nagel) following the manufacturer's instructions. DNA was eluted to 50 µl and was stored at -80°C until amplification (typically within 1-5 days).

2.3.4 Sequences for lab strains

We included sequence information for several *Synechococcus* strains that were isolated off the Scripps pier and have been maintained in the lab (at 20ºC, in F/4, Guillard and Ryther (1962), kept in continuous light and transferred approximately twice a month). DNA was

extracted by boiling cultures at 100ºC for 10 min, and adding 1 µl of this to a 25 µl PCR reaction. We amplified the ITS amplicon using the primers (F: GGATCACCTCCTAACAGGGAG, R: AGGTTAGGAGACTCGAACTC) and PCR program described in Choi et al. (2014): denaturation for 5 min at 94°C, 35 cycles of 94°C for 45 s, 50°C for 45 s, and 72°C for 90 s, and a final 10 min elongation at 72°C. PCR product was sent to Eton Biosciences (Eton Biosciences, San Diego, CA, USA) for Sanger sequencing and resulting reads were assembled using CLC Workbench v 8 (https://www.qiagenbioinformatics.com/).

2.3.5 Amplicon Sequencing

All DNA was sent to RTL Genomics (Lubbock, TX, USA) for sequencing on an Illumina MiSeq at a depth of 10K reads. Samples were sequenced according to the Research and Testing Laboratory (Lubbock, TX, USA) protocol, described in S3.

2.3.6 rpoC1 clone libraries

For some of the samples, we made clone libraries of the RNA Polymerase C (*rpoC1*) gene to compare with the ITS sequence results. Methods, samples, and results of this preliminary work are reported in S3 and S4.

2.3.7 Mock communities

In order to evaluate whether sequence results using these primers might accurately reflect the proportions of *Synechococcus* cells of various strains present in the sample, we created several mock communities from lab strains. We used a hemocytometer to enumerate *Synechococcus* cells from log-phase cultures of CC9311 (Clade I), CC9902 (Clade IV), CC9605 (Clade II), and CC9701 (Clade II). We combined these in different sets of proportions (reported in S5). These were filtered on to 0.2 µm Supor filters, stored at -80ºC, and processed for DNA sequencing in the same manner that we filtered and preserved all pier samples.

Additionally, we grew cultures of CC9902 in F/4 and low-N F/4 ($1/10^{th}$ the normal nitrate concentration). 200 µl of each culture was filtered onto 0.2 µm Supor filters and processed as stated above. Cultures were filtered when the F/4 culture was in log-phase and the low-N culture showed nitrogen limitation (S5).

2.3.8 Data analysis

ITS sequences were processed using a combination of open-source software programs. Forward and reverse reads were assembled using mothur (Schloss et al., 2009) and ASVs were assigned using the deblur workflow (Amir et al., 2017) with a trim size of 350 bp. We used mothur's classify.seqs command for classifying ASVs against a database of cyanobacterial 16S-23S ITS sequences compiled by Choi et al. (2014). Some sequence classifications were later adjusted as necessary based on the phylogenetic tree we generated (see below).

To examine the relationship of our sequences to one another and to known *Synechococcus* strains, we added ITS sequences for known marine *Synechococcus* strains from IMG-JGI (Chen et al., 2019, https://img.jgi.doe.gov/) as well as sequences from our lab strains to the amplicon sequences. This compiled fasta file was aligned and trimmed using mothur. RaxML (Stamatakis, 2014) was used to generate a maximum likelihood tree with 1000 bootstraps. This tree used a GTR+I+G model as determined using JModelTest (Darriba et al., 2012; Guindon & Gascuel, 2003) with sequences from *Synechococcus* subcluster 5.3 as an outgroup.

All other analyses were conducted using R or Python, including the R phyloseq package (Mcmurdie & Holmes, 2013).

2.4 RESULTS

2.4.1 Numerous co-existing ASVs found in the Southern California coastal ecosystem

175 ASVs were found using deblur analysis amongst all the sites. Of these, 35 were classified as clade I and 73 as clade IV, making these two clades both the most abundant and the most diverse. A maximum likelihood tree of all sequences, along with some lab strains and other cultured strains whose genomes have been sequenced, revealed a clear distinction of clades I, III, IV, and CRD1, with strong (>88%) bootstrap support (Fig. 2.1). Classification of these sequences using mothur's classify.seqs command and the Choi et al. (2014) sequence database was consistent with their position on the tree. However, sequences classified as clade II and CRD2 using the database were not monophyletic on the tree and instead formed three and two clusters respectively. ASVs that were classified as CRD2 using the database formed two separate clusters but one of these had low (<93%) identity with any CRD2 sequences within the database. We are naming this cluster "CCE Clade" as it does not appear to belong to CRD2 and forms a cluster of ASVs found exclusively at offshore sites. In the case of clade II, we believe the multiple clusters in Fig. 2.1 to be a consequence of our using a shorter length amplicon than that used to generate the original database and its classifications. We found that in a tree made with the longer, 750 bp region, these distinct clusters within clade II were maintained but formed a supported monophyletic clade together (S6). Because the positions of the clade II clusters in the tree made with the shorter region (Fig. 2.1) were poorly supported, we retain their clade identifications but refer to the clusters separately (II-A, II-B, II-C).

The maximum sequence-dissimilarity among all the ASVs was approximately 13%. Within individual clades, minimum identity was 96% (clade I), 94.9% (clade IV), 92% (clade II

all), 98% (II-A), 97.1% (II-B), 93.2% (II-C), 96.9% (CRD1), 96.6% (CRD2-A), 98.9% (CCE clade).

2.4.2 A comparison of four different blooms at the same sampling site

From 2010 to 2016, the average *Synechoocccus* abundance was 118,759 cells/ml. The *Synechococcus* population at the Scripps pier typically remains below a density of 200,000 cells/ml, with several blooms a year rising above this concentration, but with the timing and size of blooms varying greatly year to year. The maximum bloom densities also remained below 300,000 cells/ml from 2010 through 2013, and exceeded this concentration once in 2014 and twice in 2015. Most notably, there was a large bloom in summer 2016 where the density rose above 1,000,000 cells/ml (Fig. 2.2A) – a peak density greater than any detected since 2005, and up through June 2019 (Tai & Palenik, 2009, Palenik lab unpubl. data). In 2011, there were three blooms where the concentration increased from below 100,000 cells/ml to above 200,000 cells/ml and back down to the original concentrations within several weeks. In 2012, this only occurred once, though the population rose above 200,000 cells/ml two other times with a slower rate of increase. *Synechococcus* abundance had a loose but significant correlation with surface temperature (Pearson's $r = 0.31$, $p=1e-10$) but no correlation with chlorophyll. Without the warmer and larger 2016 bloom, this correlation decreases.

Synechococcus diversity at the Scripps pier was consistently dominated by clades I and IV, with the exception of the summer of 2016 during which an abnormally large bloom coincided with an atypical abundance and variety of clade II sequences (Fig. 2.2A-C). Members of other clades were occasionally observed, including a small number of sequences from CRD1 and CRD2 in summer 2011 and 2012, and a small number of sequences from clades III, XVII, and WPC1 in 2016.

The four blooms selected for comparison (Spring 2011, Summer 2011, Summer 2012, Summer 2016) differed from one another in both their initial *Synechococcus* community composition as well as in the compositional shifts that occurred over the course of the blooms (Fig. 2.2C). For example, at the start of the summer 2011 bloom, the ratio of clade I to clade IV was higher than at any other time point sampled, and decreased over the course of the 2011 bloom. In contrast, there was no clear directional change in the clade I:IV ratio in spring 2011 or summer 2012, but the ratio itself differed between the blooms (with clade IV much more abundant relative to clade I in summer 2012). We highlight this clade I to clade IV ratio because these are typically the most dominant clades at our site and this ratio has been tracked in other studies (Tai & Palenik, 2009). The trend of clade I decreasing relative to clade IV in summer 2011 was also observed in sequences from clone libraries of the *rpoC1* gene (S4). The *rpoC1* libraries from summer 2012 also revealed no shift in the clade I:IV ratio before and after the bloom. However, the summer 2012 *rpoC1* libraries contained approximately the same number of clones from clade I and clade IV whereas the ITS amplicon sequence data contained far more clade IV reads relative to clade I. Given the small number of clones sequenced, it is likely that a greater sequencing depth for *rpoC1* would more accurately reflect environmental abundances. Thus in general our ITS results are confirmed by some limited use of *rpoC1*.

The summer 2016 bloom was the warmest of the four studied, with an average surface temperature of 22.4ºC over the course of the dates sampled. Clade IV remained more abundant than clade I but over the course of the large *Synechococcus* bloom, clade II became the most abundant clade overall. On four of the 2016 sampling dates clade II comprised over 50% of all reads. Its abundance had declined by two weeks after the peak of the bloom, though it was still present and more abundant than clade I. We "quantified" counts of each clade by converting the

proportions of sequence reads to proportions of total *Synechococcus* cell abundance and found that the clade IV abundances in 2016 were still greater during the peak of the bloom (as well as on 18 April) than during any other bloom. However, the calculated density of clade II during the bloom represented a greater abundance than the total *Synechococcus* abundance during any other bloom (S7). Clade II sequences were absent entirely from all other samples except 26 July and 11 Aug 2012, where they were <1.5% of the total sequences. This is congruent with previous studies at the Scripps pier where clade II was either not detected (Tai et al., 2011) or detected at very low abundances (Tai & Palenik, 2009).

In order to determine whether we could use sequence data as a general proxy for relative abundances of different *Synechococcus* clades to one another, we examined the clade composition of sequence data obtained from five mock communities (S5). These communities consisted of different combinations of cultured *Synechococcus* strains from clades I, IV, II (two strains), and CRD1 (respectively, CC9311, CC9902, CC9701, CC9605, and CC9305). In every case, CC9311 was slightly over-represented in the sequence data relative to its known proportion of the cells added to the community (S8). While the current set of mock communities does not allow us to make any quantitative conversion factor, we can conclude that the greater proportion of clade IV sequences in our data does likely correspond with a greater abundance of clade IV cells. Importantly, the mock community containing clade II representatives had clade II sequences in approximately the same proportions as the cell composition, confirming that clade II is highly abundant during the 2016 bloom.

2.4.3 Hidden switching of within-clade variants over the course of blooms

While the *Synechococcus* community at the Scripps pier consisted mainly of clade I and IV members (with some members of CRD2 detected in 2011 and the high abundance of clade II

in 2016), there were multiple variants within each clade present at any given time (Fig. 2.3). For both clades I and IV, a single ASV was typically the most abundant representative of its clade (ASV 6 for clade I, ASV 2 for clade IV), but several other ASVs were consistently present at lower abundances.

There were observable shifts in the ratios of ASVs within a clade over the course of each bloom. For example, over the course of the spring 2011 bloom, clade IV was consistently dominated by ASV 2, but there was a shift in the dominant clade I ASV during the decline of the bloom, from ASV 6 to ASV 21. This was the only case where ASV 21 was the most abundant clade I ASV.

In the summer blooms of 2011 and 2012, the peak bloom dates (7 July 2011 and 2 Aug 2012) there was a shift from clade I ASV 6 to ASV 55 and from clade IV ASV 2 to ASV 4 just prior to the decline of the bloom. The typically dominant clade I and IV ASVs (respectively, ASVs 6 and 2) returned to this role for the decline of both of these blooms. It is also noteworthy that, during the peak of the 2016 bloom (11 Aug 2016), the two main clade I and IV ASVs at the pier (ASV 6 and 2 respectively) were completely absent, though ASVs 55 and 4 were present. Their sequences were detected on all other dates, including the rest of the 2016 dates sampled. We also observed a similar shift within clade II members over the course of the 2016 bloom. ASVs 119 (clade II-C) and 120 (clade II-A) were dominant before and after the bloom with a shift to ASVs 142 (clade II-C) and 166 (clade II-A) on the peak bloom date

The rapid temporary switching between the more ubiquitous ASV and the secondary ASV that became abundant during the peak of the bloom was consistently between pairs of ASVs that differed in the 22nd position of the amplicon sequence. In every case, the more ubiquitous variant (ASV 6 for clade I, ASV 2 for clade IV, ASV 120 for clade II-A, and ASV

119 for clade II-B) had a "G" in the 22nd position while the more transient variant had an "A" (ASV 55 for clade I, ASV 4 for clade IV, ASV 166 for clade II-A, and ASV 142 for clade II-C). This was the only base change between these pairs of ASVs (Fig. 2.4).

Because of the unusually consistent nature of this ASV-switching (with the same type of switch occurring in multiple clades), it was unclear whether this actually represented "microdiversity" in the sense of a separate population becoming temporarily more abundant. To investigate whether this phenomenon could instead be a physiological response to environmental conditions, we grew cultures of *Synechococcus* CC9902 in either regular or low-nitrate F/4 media. We then sequenced the ITS region as done with our environmental samples to see whether there would be multiple ASVs within a single sample, and whether both samples would have the same ASV composition. We found that the sequences from the control culture contained both G- and A- variants (i.e. sequences corresponding with both ASV 2 and ASV 4, respectively). Approximately 18% of sequences had an A in the 22nd position and the rest had a G. In the low-N culture, 100% of the sequences belonged to ASV 2, with a G in the 22nd position (Fig. 2.4).

2.4.4 Spatial variation in Synechococcus community composition

We observed variation in the *Synechococcus* clades present at different cycles (sites) in the CCE (Fig. 2.5), with greater ASV richness further offshore. Clades II, CRD2 and WPC1 were only present at the oligotrophic, offshore cycles (1 and 2), and clades CRD1, and XVII were present at cycles 1, 2 and 3. However, only clades I and IV were found at the nearshore cycle 4 (with the exception of three CRD1 reads). Within individual clades, some ASVs were more abundant offshore (ASVs 134 & ASV 10 in clade I, ASVs 7 and 74 in clade IV, ASV 87 in

CRD1) while others were observed only at nearshore sites (ASV 21 in clade I and ASV 4 in clade IV) (Fig. 2.3).

Synechococcus depth profiles differed at each sampling site, with cell numbers peaking near the deep chlorophyll maxima in each case (S2). At the two sites furthest offshore (cycle 1, the offshore stratified region, and cycle 2, the core of the California Current proper), clade IV was more abundant relative to clade I at all depths except 150 m (Fig. 2.5). In cycle 3, the wind stress curl upwelling domain, clade IV was more abundant than clade I at 5, 12, and 30 m, but clade I became more abundant than clade IV at 60 m and 150 m. In the coastal boundary upwelling region (cycle 4), Clade I was more abundant relative to clade IV at most depths. The deepest samples (150 m) always comprised mostly of clade I, though *Synechococcus* density was much lower at this depth (S 2.6.2). Thus there is now substantial evidence that clade I becomes dominant relative to clade IV with increasing depth.

There were 52 ASVs shared between the pier and CCE sampling sites; the rest were geographically distinct (Fig. 2.6A). The most abundant clade I and IV ASVs from the pier (ASVs 6 and 2 respectively) were also present and abundant at all CCE sampling sites above 90 meters; furthermore, ASV 6 was the most abundant ASV in every single deep (150 m) sample. ASV 7, a presumably oligotrophic specialist member of clade IV, was detected at the pier but always at low levels, but most ASVs belonging to "offshore" clades as identified above were not found at the pier. While members of clade II were found at the pier and at the two furthest offshore cycles, most (28 of 30) clade II ASVs were specific either to the pier or to the CCE sites, despite the fact that most of the clade II sequences at the pier were found in 2016, just months after the CCE sampling occurred. Overall, the pier and CCE *Synechococcus*

communities were fairly distinct, and the pier communities during 2011 and 2012 particularly clustered separately from that in 2016 (Fig. 2.6B).

2.5 DISCUSSION

It is well known that *Synechococcus* is a diverse genus with numerous evolutionary adaptations allowing it to be widespread among the world's oceans. However, as this sequencing effort demonstrates, its true diversity remains under-characterized. Lee et al. (2019) found that the majority of environmental sequences are not identical to reference genomes. Indeed in the case of clade I the most abundant variant in our data did not share the same ITS sequence as its closest cultured genome representative, CC9311 (although in the case of clade IV, CC9902 corresponded with ASV2 which was generally the most abundant clade IV ASV). Additionally, our results add another layer to the question of why different clades of *Synechococcus* co-occur in the same environment; we found that even closely related variants within the same *Synechococcus* clade can co-occur at a single site with detectable fluctuations in their abundance over fairly short time scales. Furthermore, we observed in multiple cases a strong reduction in the abundance of an ASV between one week and the next, with a corresponding appearance or large increase of a previously undetected or sparse ASV, with the two switching ASVs differing by as little as a single base pair in the ITS region sequenced. Given that these can still be considered sequence variants, we will henceforth continue to refer to them as ASVs, but we suggest that these may not actually reflect distinct taxa as the ASV designation typically implies. Ultimately, we believe our sequence results reveal both shifts in microdiverse sub-populations as well as a potential physiological response in the dominant ASVs captured by our sequencing approach.

Understanding the drivers behind the initiation of phytoplankton blooms is of great interest for numerous reasons, including the evaluation of ecological hypotheses and potential for prediction of future bloom events. At the Scripps pier site we know from over a decade of monitoring that blooms of *Synechococcus* typically occur a few times a year and last on the order of days to weeks. Because we don't observe an explicit seasonal abundance pattern, our site is well suited for studying the drivers of both bloom onset and decline. While *Synechococcus* abundance does have a loose correlation with water temperature, no strong and consistent patterns with other biotic or abiotic factors have been confirmed at our site to date. We have previously used sequencing methods to identify putative *Synechococcus* grazers based on their increased abundance during or immediately following *Synechococcus* blooms (Nagarkar et al., 2018), but particular biological interactions have not yet been corroborated as consistent driving forces of initiation or termination of these blooms. Two prominent hypotheses for biotic factors contributing to bloom decline are grazing and viral lysis (Landry & Hassett, 1982; Proctor & Fuhrman, 1990), while changes in environmental conditions could also affect the success or mortality of *Synechococcus*. All of these factors have the potential to act with specificity on the sub-clade scale.

While *Synechococcus* blooms can readily be detected and tracked using microscopy or flow cytometry, these methods cannot distinguish between different clades of *Synechococcus*. However, studies at our and other sites have found that diverse *Synechococcus* variants, at the clade and sub-clade level, are often present at a single location, and furthermore that the within-*Synechococcus* diversity can shift drastically over small spatial or temporal scales (Mackey et al., 2017; Paerl et al., 2011; Slack et al., 2014; Tai et al., 2011). In this study we found both cases where the *Synechococcus* community composition at the clade level changes considerably over

the course of the bloom (summer 2011, summer 2016) and other cases where it remains fairly consistent (spring 2011, summer 2012). Some common *Synechococcus* grazers, including *Paraphysmonas sp.* and *Pteridomonas sp.* have been demonstrated to have high prey selectivity, with differential growth success on *Synechococcus* strains even within the same clade (Zwirglmaier et al., 2009). A shift in clade composition seen over the course of a bloom as in summer 2011 could be indicative of grazers targeting one clade over the other.

The 2016 bloom was highly atypical both in its high abundances of *Synechococcus* and in the appearance of clade II sequences. Among all *Synechococcus* counts measured between 2010 and 2016, this represented the largest bloom both in terms of *Synechococcus* abundance at its peak as well as the fastest increase in *Synechococcus* density between one sampling point and the next. The prominent presence of clade II was also unique; clade II is typically found in the open ocean, particularly in warmer and more oligotrophic (especially low-phosphate) environments, and has been found to co-occur with clades III and X (Sohm et al., 2016). We did detect a small number of clade III reads at the pier site during the same time as the 2016 bloom, though no sequences were classified as clade X. Additionally, the clade II ASVs detected that were unique to that bloom and not found at the pier in other years or at any of our offshore sites despite the cruise having been only a few months prior, in April 2016. Conversely, only two clade II ASVs from the CCE sites were detected at the pier, at very low abundances and only in 2012. Two atypical warming events preceded the 2016 bloom: a series of warm anomalies in the North Pacific along with an El Niño that persisted through spring 2016 (Di Lorenzo & Mantua, 2016). This time period was marked by features including reduced frequency of oceanic fronts (Kahru et al., 2018). It is unclear whether this anomalous bloom represents the arrival of a distinct water

parcel to the coast or is a result of conditions shifting to be favorable for members of clade II. It is possible that this may be a more common occurrence in the future with climate warming.

Geographical niche partitioning has been extensively documented in *Synechococcus* and is hypothesized to indicate adaptations of ecotypes to different temperature, light and nutrient regimes, as well as environments where these resources are consistently available compared to highly variable resources as might be found on the coast (Zwirglmaier et al., 2007). From prior studies it is well established that clades I and IV are found to co-occur in coastal environments, and members of these clades have consistently been dominant at the Scripps Pier (Tai & Palenik, 2009). However, they were also the most abundant members of all offshore samples including the more oligotrophic sites. The dominance of clade I at depth compared to clade IV seen here brings up the hypothesis that the observed seasonal pattern in the clade I:IV ratio seen in surface waters at the SIO pier (Tai et al., 2011) is due to seasonal stratification which would favor clade IV at the surface and clade I at depth and thus decrease the I:IV ratio.

On the ASV level, the main pier clade I ASV (ASV 6) was surprisingly also the most abundant member of clade I at most offshore sites, and the most abundant ASV overall in each deep sample (150 m). This finding is consistent with trends seen in 2011, where clade I was found at the pier and increased with depth in offshore mesotrophic waters (Tai et al., 2011). It appears that this particular clade I ASV is more of a generalist, capable of adapting to different light, nutrient, and temperature regimes. There is a small amount of evidence that Clade I ASVs 1, 10, and 134 are oligotrophic variants. At the pier Clade IV ASV 2 was by far the most abundant and ASV 7 was not even detected in some years. However, offshore there was co-dominance of ASVs 2 and 7 . ASV 7 is likely a preferentially oligotrophic clade IV variant. Thus we have found further evidence of microdiversity in *Synechococcus* having a

biogeographical pattern. We also acknowledge that ITS as a marker may not be sufficient to reveal all the "true" microdiversity within a sample.

An important consideration for studies that utilize environmental sequencing is whether the sequence-based characterization of the community accurately reflects relative abundances of its members. In order to examine this we sequenced several mock communities created from combinations of lab cultures. While the sequence results from five cultured strains cannot be generalized to the entire set of ASVs found in our environmental samples, they offer a preliminary confirmation that sequencing of this amplicon can adequately represent environmental abundances. We found clade IV sequences to be more abundant than clade I sequences at the majority of our time points and sites and believe the sequence data reflects the relative abundance of those clades because of results from sequencing mock communities. If anything, the mock community results, in which clade I sequences are consistently slightly overrepresented, suggest that clade IV might be even more abundant environmentally than the sequence data indicate. This skew in the relative proportions of resulting sequence data is likely due to differing ease of DNA extraction or PCR primer bias between the different strains, as the cultures used are known to have the same copy number for the ITS region sequenced. Importantly, some of the mock communities contained two different clade I or IV ASVs, a result discussed further below.

Our sequence data indicate that *Synechococcus* blooms consist of several sequence variants with individual temporal patterns. We were able to detect large changes in the sub-clade-level community composition, with both the dominant clades (I and IV) experiencing a spike in a less abundant variant during the peak of three separate blooms (the exception was spring 2011). A similar phenomenon was also found with clade II ASVs during the 2016 bloom.

These patterns are remarkable in their consistency over multiple blooms and between different clades. In each case the "switching" ASVs show the same G to A shift in the 22nd position of the amplicon sequenced, and in some of the blooms this switch occurs between multiple pairs of ASVs from different clades. To our knowledge no shift in sequence has been described at such fine resolution, both in terms of being a single nucleotide switch and in being on the scale of days. We also were unable to find *Synechococcus* sequences with a "G" in that position in several databases, including the Choi et al. (2014) database we used for classification, and the sequences deposited in several other *Synechococcus* sequencing efforts (Lee et al., 2019; Mazard et al., 2011). However, amplicon sequences deposited in the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) by Choi et al. (2014, 2015) did include variants with this genotype.

Interestingly, the sequences of both CC9311 and CC9902 from our mock communities were dominated by their respective G-variants. In the case of CC9902, we found that cultures of *Synechococcus* CC9902 grown in standard F/4 media contained both sequence variants, with the A-variant accounting for approximately 18% of sequence reads, but when nitrogen-limited, only the G-variant appeared. This indicates that the variant-switching we observe over the course of blooms is likely a physiological response rather than a shift in microdiversity. Unfortunately, nutrient data was not available on all bloom dates sampled and thus we cannot describe changes in nutrient conditions over the course of each bloom. One possibility is that *Synechococcus* undergoes DNA modifications such methylation or oxidation in response to its environment and these affect correct base pair calling during amplification and sequencing. The cyanobacterium *Synechocystis* has been found to respond to nitrogen starvation with changes in methylation (Hu et al., 2018). And while methylation is not known to affect Illumina readouts, one example of a

DNA modification that did induce a sequencing artifact was described in Costello et al (2013), where the acoustic DNA shearing protocol was found to induce guanine oxidation.

Recent characterizations of *Synechococcus* using sequencing have indicated that a genus already known to be diverse consists of far more species and sub-species level variants than previously realized. The fine-scale spatial and temporal specialization of *Synechococcus* sub-types demonstrates that combinations of environmental factors can create small and transient niches that individual populations readily take advantage of. Here we have demonstrated that this microdiversity has distinct patterns with geography and depth and can shift drastically on the scale of days. Furthermore, we report a previously undescribed phenomenon – switching between sequence variants that appears to be due to a physiological response rather than a population shift.

2.6 FIGURES

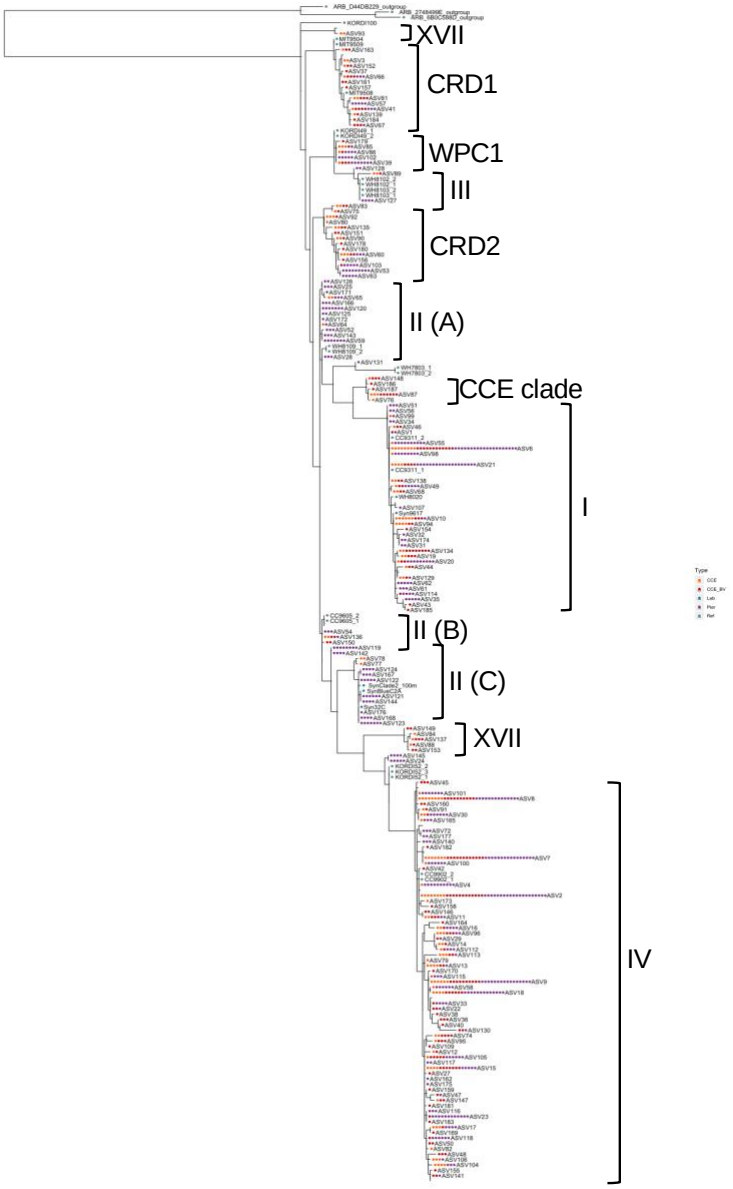


Figure 2.1. Maximum likelihood tree of *Synechococcus* OTUs along with selected cultured samples and sequences from databases along with clade designations used for ASVs hereafter.. The number of circles by each OTU represents the number of samples it was found in, and circle color represents the type of sample. Three sequences from *Synechococcus* 5.3 used as outgroups.

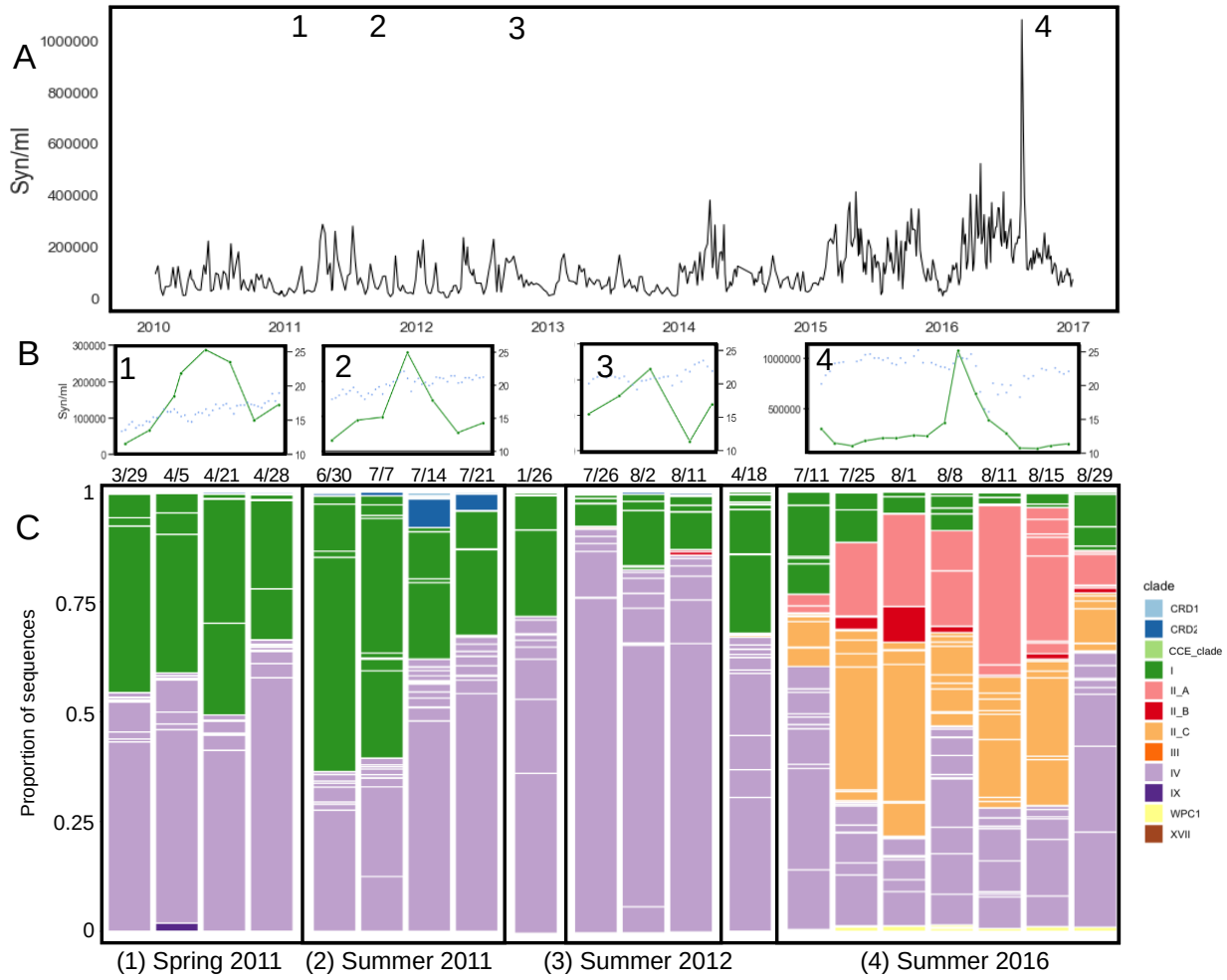


Figure 2.2. Comparison of four *Synechococcus* blooms at the Scripps pier. (A) *Synechococcus* density over the entire time period between 2010 – 2016, measured weekly or bi-weekly. (B) *Synechococcus* density at the SIO pier during the specific blooms (green) and surface seawater temperature (blue). (C) Relative abundances of different *Synechococcus* clades during selected time points over the course of *Synechococcus* blooms. White dividing lines indicate ASVs falling within the colored clades.

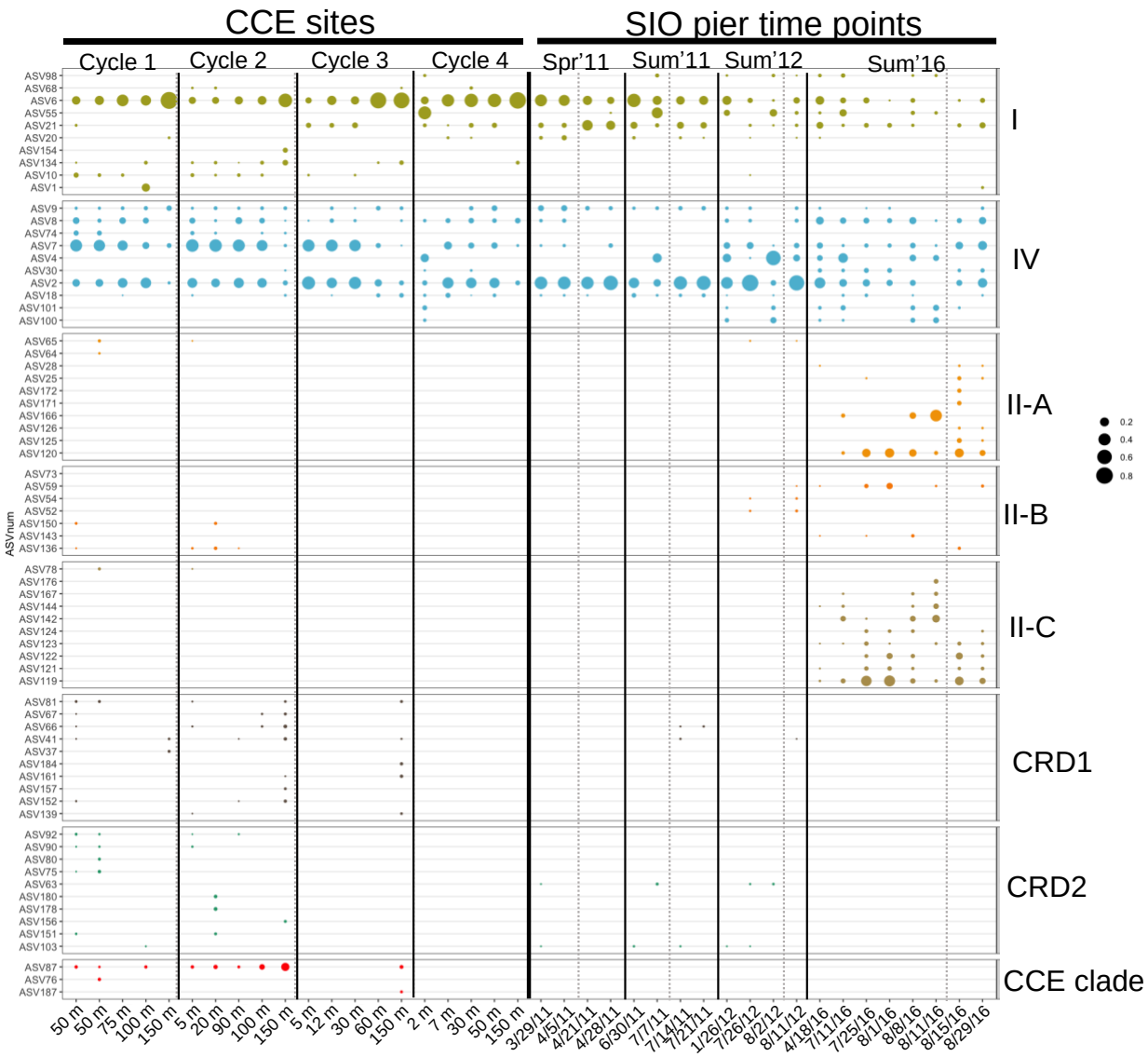


Figure 2.3. Relative abundance of *Synechococcus* ASVs among pier and CCE samples. Circle size of an ASV indicates the proportion of the total sample its sequences comprise. Dark vertical lines separate the different cycles or the different blooms that were evaluated. Grey dotted lines delineate the termination of the bloom (in each case, time points before the dotted line represent the peak abundance measured during that bloom and time points after dotted line are lower than the peak abundance). Only the top ten most abundant ASVs are shown for each clade (or fewer if there were not ten).

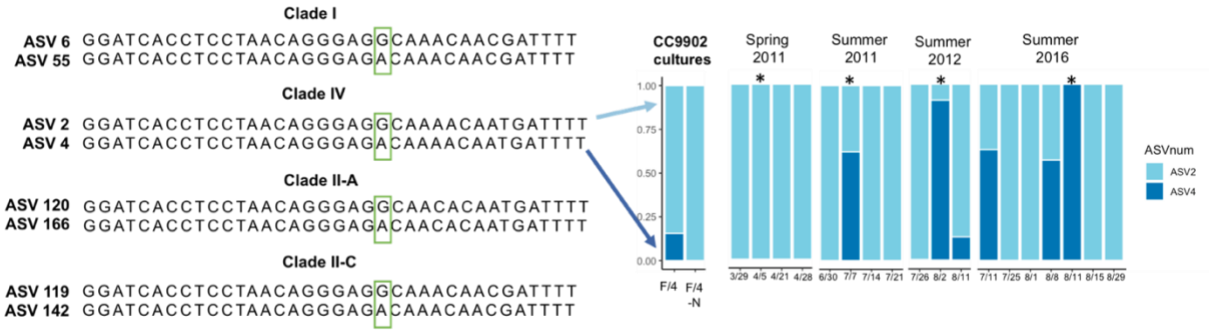


Figure 2.4. Left: comparative alignments of the beginning stretch of the ITS amplicon between pairs of temporally-switching ASVs within four clades. All four pairs differ in the same position. In each case, the rest of the sequence shown is identical between each pair. Right: relative abundances of just the G-variant (ASV 2) and A-variant (ASV 4) in *Synechococcus* CC9902 cultures grown in regular F/4 media or nitrogen-poor F/4 media and in the pier samples. Stars denote the peaks of each bloom.

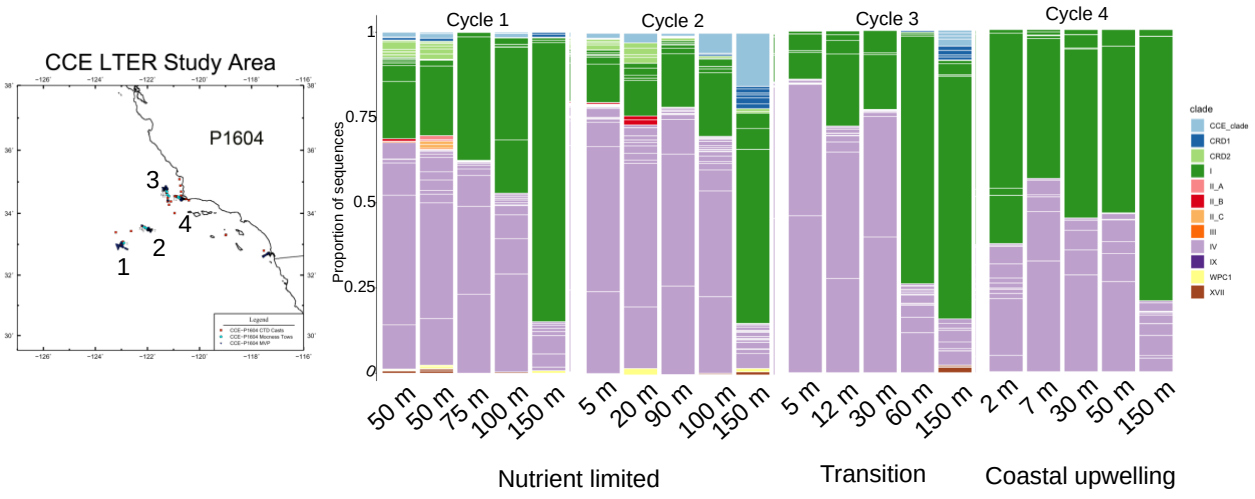


Figure 2.5. Map of sampling sites (cycles) from CCE P1604 cruise and *Synechococcus* community composition at multiple depths within each cycle.

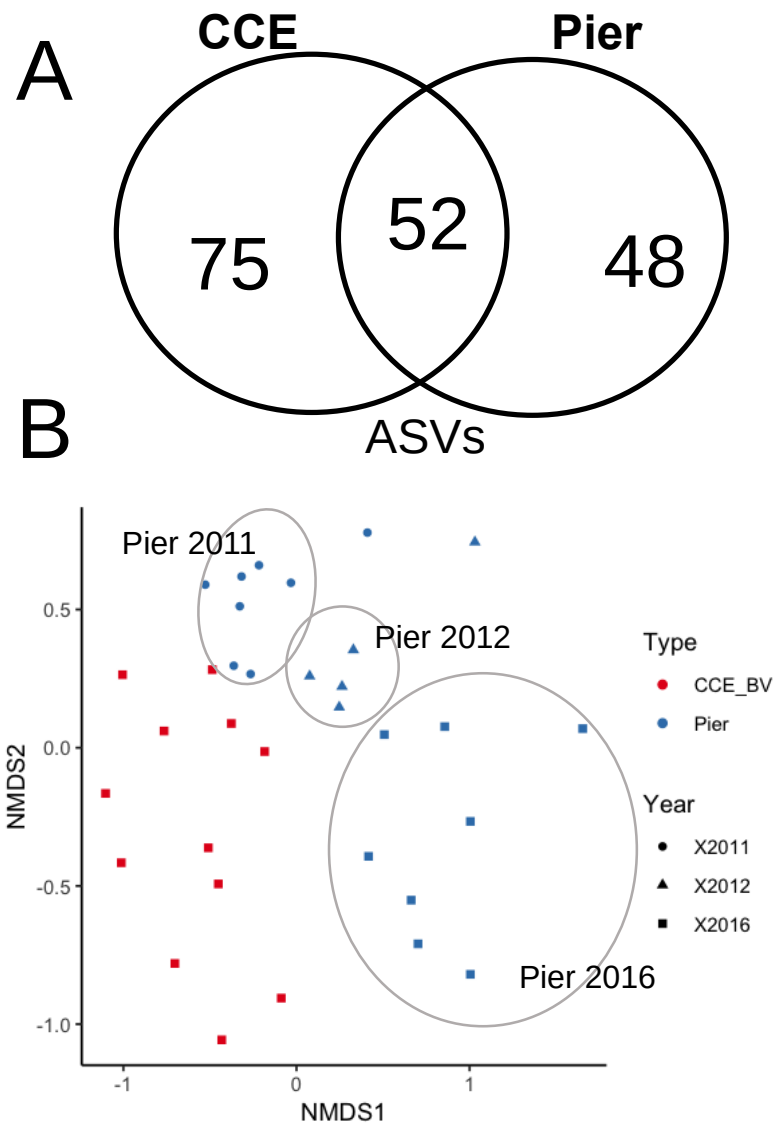


Figure 2.6. (A) Number of ASVs unique to, and shared among, the SIO pier and offshore samples. (B) NMDS plot of Bray-Curtis dissimilarity between all samples with colors designating sample type and shape designating year. Grey circles drawn to highlight 2011, 2012, and 2016 pier samples.

2.7 ACKNOWLEDGEMENTS

Funding for this project was provided through a National Science Foundation grant (DEB-1233085) to Brian Palenik and National Science Foundation grant OCE-1614359 to the CCE-LTER site. Analyses of molecular samples was partly funded by the Graduate Student Excellence Research Award of the Scripps Institution of Oceanography to B. Valencia. B. Valencia was also supported by a scholarship (529-2011) from the Colombian Administrative Department of Science, Technology and Innovation (COLCIENCIAS).

We thank Mike Landry his valuable input and for providing FCM data from the P1604 CCE cruise and Traci Yuen for her contributions in making some of the preliminary clone libraries. Scripps Pier measurements collected by the Birch Aquarium at Scripps staff and volunteers. Data provided by the Shore Stations Program sponsored at Scripps Institution of Oceanography by California State Parks, Division of Boating and Waterways. Contact: shorestation@ucsd.edu. We thank Melissa Carter, Kristi Seech, and James Fumo for their consistent help and patience with pier sampling.

Chapter 2, in full, is currently being prepared for submission for publication of the material with the following co-authors: Nagarkar, M., Wang, M., Valencia-Ramirez, B., Palenik, B. The dissertation author was the primary author of this paper.

2.8 REFERENCES

Acinas, S. G., Klepac-ceraj, V., & Hunt, D. E. (2003). Fine-scale phylogenetic architecture of a complex bacterial community. *Nature, 430(6999), 551*.

Agawin, N. S. R., Duarte, C. M., & Agustí, S. (1998). Growth and abundance of Synechococcus sp. in a Mediterranean Bay:seasonality and relationship with temperature. *Marine Ecology Progress Series*, *170*, 45–53. https://doi.org/10.3354/meps170045

Agawin, N. S. R., Duarte, C. M., & Agustí, S. (2000). Nutrient and temperature control of the contribution of picoplankton to phytoplankton biomass and production. *Limnology and Oceanography*, *45*(3), 591–600. https://doi.org/10.4319/lo.2000.45.3.0591

Ahlgren, N. A., & Rocap, G. (2012). Diversity and distribution of marine Synechococcus : multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Frontiers in Microbiology*, *3*, 1–24. https://doi.org/10.3389/fmicb.2012.00213

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., et al. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *MSystems*, *2*(2), 1–7.

Chen, I. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., et al. (2019). IMG/Mv. 5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Research*, *47*(October 2018), 666–677. https://doi.org/10.1093/nar/gky901

Choi, D. H., & Noh, J. H. (2009). Phylogenetic diversity of Synechococcus strains isolated from the East China Sea and the East Sea. *FEMS Microbiology Ecology*, *69*, 439–448. https://doi.org/10.1111/j.1574-6941.2009.00729.x

Choi, D. H., Hoon Noh, J., & Lee, J.-H. (2014). Application of Pyrosequencing Method for Investigating the Diversity of Synechococcus Subcluster 5.1 in Open Ocean. *Microbes and Environments*, *29*(1), 17–22. https://doi.org/10.1264/jsme2.ME13063

Collado-Fabbri, S., Vaulot, D., & Ulloa, O. (2011). Structure and seasonal dynamics of the eukaryotic picophytoplankton community in a wind-driven coastal upwelling ecosystem. *Limnology and Oceanography*, *56*(6), 2334–2346. https://doi.org/10.4319/lo.2011.56.6.2334

Collier, J. L., Palenik, B., Olson, R. J., Chisholm, S. W., Zettler, E. R., Armbrust, E. V., et al. (2003). Phycoerythrin-containing picoplankton in the Southern California Bight. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *50*(14–16), 2405–2422. https://doi.org/10.1016/S0967-0645(03)00127-9

Costello, M., Pugh, T. J., Fennell, T. J., Stewart, C., Lichtenstein, L., Meldrim, J. C., et al. (2013). Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Research*, *41*(6), 1–12. https://doi.org/10.1093/nar/gks1443

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2 : more models, new heuristics and parallel computing CircadiOmics: integrating circadian genomics, transcriptomics, proteomics. *Nature Methods*, *9*(8), 772. https://doi.org/10.1038/nmeth.2109

Farrant, G. K., Doré, H., Cornejo-castillo, F. M., Partensky, F., Ratin, M., & Garczarek, L. (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *PNAS, 113(24), E3365-E3374*. https://doi.org/10.1073/pnas.1524865113

Ferris, M. J., & Palenik, B. (1998). Niche adaptation in ocean cyanobacteria. *Nature*, *396*, 226–228.

Glover, H. E., Prezelin, B., Campbell, L., Wyman, M., & Garside, C. (1988). A nitrate-dependent Synechococcus bloom in surface Sargasso Sea water. *Nature*, *331*(January), 161–163. https://doi.org/10.1038/331161a0

Guillard, R. R. L., & Ryther, J. H. (1962). Studies of marine planktonic diatoms: I. *Cyclotella nana* Husttedt, and *Detonula confervacea* (Cleve) Gran. *Canadian Journal of Mirobiology*, (229–239).

Guindon, S., & Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies. Systematic biology, 52(5), 696-704. https://doi.org/10.1080/10635150390235520

Hu, L., Xiao, P., Jiang, Y., Dong, M., Chen, Z., Li, H., et al. (2018). Transgenerational epigenetic inheritance under environmental stress by genome-wide DNA methylation profiling in cyanobacterium. *Frontiers in Microbiology*, *9*, 1–11. https://doi.org/10.3389/fmicb.2018.01479

Hunter-Cevera, K. R., Post, A. F., Peacock, E. E., & Sosik, H. M. (2016). Diversity of Synechococcus at the Martha's Vineyard Coastal Observatory : Insights from Culture Isolations, Clone Libraries, and Flow Cytometry. *Microbial Ecology*, *71*, 276–289. https://doi.org/10.1007/s00248-015-0644-1

Kahru, M., Jacox, M. G., & Ohman, M. D. (2018). CCE1: Decrease in the frequency of oceanic fronts and surface chlorophyll concentration in the California Current System during the 2014–2016 northeast Pacific warm anomalies. *Deep-Sea Research Part I: Oceanographic Research Papers*, *140*, 4–13. https://doi.org/10.1016/j.dsr.2018.04.007

Kashtan, N., Roggensack, S. E., Rodrigue, S., Thompson, J. W., Biller, S. J., Coe, A., et al. (2014). Single-Cell Genomics Reveals Hundreds of Coexisting Subpopulations in Wild Prochlorococcus. *Science, 344(6182), 416-420.*

Landry, M. R., & Hassett, R. P. (1982). Estimating the grazing impact of marine micro-zooplancton. *Marine Biology*, *67*(3), 283–288. https://doi.org/10.1007/BF00397668

Larkin, A. A., & Martiny, A. C. (2017). Minireview Microdiversity shapes the traits , niche space, and biogeography of microbial taxa. *Environmental microbiology reports*, *9(2)*, *55-70*. https://doi.org/10.1111/1758-2229.12523

Lavin, P., González, B., Santibáñez, J. F., Scanlan, D. J., & Ulloa, O. (2010). Novel lineages of Prochlorococcus thrive within. *Environmental Microbiology Reports*, *2*(6), 728–738. https://doi.org/10.1111/j.1758-2229.2010.00167.x

Lee, M. D., Ahlgren, N. A., Kling, J. D., Walworth, N. G., Rocap, G., Saito, M. A., et al. (2019). Marine Synechococcus isolates representing globally abundant genomic lineages demonstrate a unique evolutionary path of genome reduction without a decrease in GC content. *Environmental Microbiology*, *00*, 1–10. https://doi.org/10.1111/1462-2920.14552

Di Lorenzo, E., & Mantua, N. (2016). Multi-year persistence of the 2014/15 North Pacific marine heatwave. *Nature Climate Change*, *6*(11), 1042–1047. https://doi.org/10.1038/nclimate3082

Mackey, K. R. M., Hunter-Cevera, K. R., Britten, G. L., Murphy, L. G., Sogin, M. L., & Huber, J. A. (2017). Seasonal Succession and Spatial Patterns of Synechococcus Microdiversity in a Salt Marsh Estuary Revealed through 16S rRNA Gene Oligotyping. *Frontiers in Microbiology*, *8*, *1496*. https://doi.org/10.3389/fmicb.2017.01496

Mazard, S., Ostrowski, M., Partensky, F., & Scanlan, D. J. (2011). Multi-locus sequence analysis, taxonomic resolution and biogeography of marine Synechococcus. *Environmental Microbiology*, *14(2)*, *372-386.* https://doi.org/10.1111/j.1462-2920.2011.02514.x

Mcmurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061217

Nagarkar, M., Countway, P. D., Du, Y., Daniels, E., Poulton, N. J., & Palenik, B. (2018). Temporal dynamics of eukaryotic microbial diversity at a coastal Pacific site. *The ISME Journal*, 2278–2291. https://doi.org/10.1038/s41396-018-0172-3

Olson, R. J., Vaulot, D., & Chisholm, S. W. (1985). Marine phytoplankton distributions measured using shipboard flow cytometry. *Deep Sea Research Part A. Oceanographic Research Papers*, *32*(10), 1273–1280.

Paerl, R. W., Johnson, K. S., Welsh, R. M., Worden, A. Z., Chavez, F. P., & Zehr, J. P. (2011). Differential distributions of Synechococcus subgroups across the California current system. Frontiers in microbiology, 2, 59.https://doi.org/10.3389/fmicb.2011.00059

Palenik, B. (1994). *Cyanobacterial Community Structure as Seen from RNA Polymerase Gene Sequence Analysis*. *Applied and Environmental Microbiology,* 60(9), 3212-3219.

Palenik, B., & Haselkorn, R. (1992). Multiple evolutionary origins of prochlorophytes, the chlorophyll b-containing prokaryotes. *Nature*, *355*, 265–267.

Patterson, J. (1998). Annual average abundance of heterotrophic bacteria and Synechococcus in surface ocean waters. *Limnology and Oceanography*.

Proctor, L. M., & Fuhrman, J. a. (1990). Viral mortality of marine bacteria and cyanobacteria. *Nature*, *343*, 60–62. https://doi.org/10.1038/343060a0

Rajaneesh, K. M., & Mitbavkar, S. (2013). Factors controlling the temporal and spatial variations in Synechococcus abundance in a monsoonal estuary. *Marine Environmental Research*, *92*, 133–143. https://doi.org/10.1016/j.marenvres.2013.09.010

Robidart, J. C., Preston, C. M., Paerl, R. W., Turk, K. A., Mosier, A. C., Francis, C. A., et al. (2011). Seasonal Synechococcus and Thaumarchaeal population dynamics examined with high resolution with remote in situ instrumentation. *The ISME Journal*, *6*(3), 513–523. https://doi.org/10.1038/ismej.2011.127

Rocap, G., Distel, D. L., Waterbury, J. B., & Chisholm, S. W. (2002). Resolution of Prochlorococcus and Synechococcus Ecotypes by Using 16S-23S Ribosomal DNA Internal Transcribed Spacer Sequences. *Applied and Environmental Microbiology*, *68*(3), 1180–1191. https://doi.org/10.1128/AEM.68.3.1180-1191.2002

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur : Open-Source , Platform-Independent , Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. https://doi.org/10.1128/AEM.01541-09

Selph, K. E., Landry, M. R., Taylor, A. G., Yang, E., Measures, C. I., Yang, J., et al. (2011). Spatially-resolved taxon-specific phytoplankton production and grazing dynamics in relation to iron distributions in the Equatorial Pacific between 110 and 140 1 W. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *58*, 358–377. https://doi.org/10.1016/j.dsr2.2010.08.014

Slack, G., Daniels, E. F., Selph, K. E., Palenik, B., & Landry, M. R. (2014). Fine spatial structure of genetically distinct picocyanobacterial populations across environmental gradients in the Costa Rica Dome. *Limnology and Oceanography, 59*(3), 705–723. https://doi.org/10.4319/lo.2014.59.3.0705

Sohm, J. A., Ahlgren, N. A., Thomson, Z. J., Williams, C., Moffett, J. W., Saito, M. A., et al. (2016). Co-occurring Synechococcus ecotypes occupy four major oceanic regimes defined by temperature , macronutrients and iron, *The ISME Journal*, *10*(2), 333–345. https://doi.org/10.1038/ismej.2015.115

Stamatakis, A. (2014). RAxML version 8 : a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stuart, R. K., Brahamsha, B., Busby, K., & Palenik, B. (2013). Genomic island genes in a coastal marine Synechococcus strain confer enhanced tolerance to copper and oxidative stress. *The ISME Journal*, *7*(6), 1139–1149. https://doi.org/10.1038/ismej.2012.175

Tai, V., & Palenik, B. (2009). Temporal variation of Synechococcus clades at a coastal Pacific Ocean monitoring site. *The ISME Journal*, *3*(8), 903–915. https://doi.org/10.1038/ismej.2009.35

Tai, V., Burton, R. S., & Palenik, B. (2011). Temporal and spatial distributions of marine Synechococcus in the Southern California Bight assessed by hybridization to bead-arrays. *Marine Ecology Progress Series*, 133–147. https://doi.org/10.3354/meps09030

Toledo, G., Palenik, B., & Brahamsha, B. (1999). Swimming Marine Synechococcus Strains with Widely Different Photosynthetic Pigment Ratios Form a Monophyletic Group. *Applied and Environmental Microbiology*, 65(12), 5247-5251.

Waterbury, J. B., Watson, S. W., Guillard, R. R. L., & Brand, L. E. (1979). Widespread occurance of a unicellular, marine, planktonic, cyanobacterium. *Nature*.

Xia, X., Vidyarathna, N. K., Palenik, B., Lee, P., Liu, H., & Diego, S. (2015). Comparison of the seasonal variation of Synechococcus assemblage structure in estuarine waters and coastal waters of Hong Kong. *Applied and Environmental Microbiology*, 81.21https://doi.org/10.1128/AEM.01895-15

Zeidner, G., Preston, C. M., Delong, E. F., Massana, R., Post, A. F., Scanlan, D. J., & Béjà, O. (2003). Molecular diversity among marine picophytoplankton as revealed by psbA analyses. *Environmental Microbiology*, *5*, 212–216.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., et al. (2007). Global phylogeography of marine Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among oceanic biomes. *Environmental Microbiology*, https://doi.org/10.1111/j.1462-2920.2007.01440.x

Zwirglmaier, K., Spence, E., Zubkov, M. V., Scanlan, D. J., & Mann, N. H. (2009). Differential grazing of two heterotrophic nanoflagellates on marine Synechococcus strains. *Environmental Microbiology*, *11*(7), 1767–1776. https://doi.org/10.1111/j.1462-2920.2009.01902.x

APPENDIX

Supplemental table S1. Synechococcus cell counts (determined by flow cytometry) and seawater temperature for all pier sampling dates that were sequenced. Surface and bottom temperature obtained from SCCOOS manual shore station data (https://scripps.ucsd.edu/programs/shorestations/shore-stations-data/). NA = data not available; * = Value not available in manual shore station data; reported automated shore station temperature.

Date	Synechococcus count (cells/ml)	Surface temperature (°C)	Bottom temperature (°C)
29 March 2011	66711.4286	14.8	14.6
5 April 2011	159734.286	16.7	16.1
21 April 2011	253797.143	16.8	16.6
28 April 2011	93922.8571	17.3	17.2
30 June 2011	97525.7143	18.8	16.7
7 July 2011	281097.143	21.1	19.6
14 July 2011	145862.857	20.3	19.7
21 July 2011	53594.2857	20.4	15.7
26 January 2012	58980	15	14.9
26 July 2012	154230.476	20.9	20.6
2 August 2012	230161.905	20.8	19.4
11 August 2012	26527.619	21.9	NA
18 April 2016	523815.238	19.9	18.3
11 July 2016	308502.857	20.3	NA
25 July 2016	214605.7143	23.9	23.3
1 August 2016	241302.8571	24.5	23.2
8 August 2016	367388.5714	22.8	22.7
11 August 2016	1081217.143	24.4	23.2
15 August 2016	651640	23	18
29 August 2016	111222.857	20.5*	NA

Supplemental table S2. Information about samples from CCE LTER P1604 Cruise. Additional metadata, including temperature, salinity, density, and dissolved nutrient concentrations, is available online (https://cce.lternet.edu/data).
NA = data not available

Cycle	Distance from shore	Depths sampled	Average Syn/ml	Average Temperature (ºC)
Cycle 1	278 km	50 m	21,007	15.9
		75 m	15,947	14.4
		100 m	NA	12.4
		150 m	NA	9.3
Cycle 2	174 km	5 m	9,161	15.2
		20 m	10,828	15.2
		90 m	17,903	14.3
		100 m	5,414	13.2
		150 m	NA	9.5
Cycle 3	55 km	5 m	32,069	13.6
		12 m	33,101	13.4
		30 m	41,484	12.9
		60 m	15,173	10.8
		150 m	NA	8.7
Cycle 4	18 km	2 m	162,215	14.1
		7 m	NA	14.4
		30 m	19,212	12.0
		50 m	NA	11.2
		150 m	NA	8.8

Supplemental text S3. Additional methods (amplicon sequencing and *rpoC1* clone library preparation).

Amplicon sequencing at RTL Genomics

Samples were amplified for sequencing in a two-step process. The forward primer was constructed with (5'-3') the Illumina i5 sequencing primer (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) and the ITS-af-fusion primer (GGATCACCTCCTAACAGGGAG) (Lavin et al., 2010). The reverse primer was constructed with (5'-3') the Illumina i7 sequencing primer GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) and the Syn-ar-fusion primer (AGGTTAGGAGACTCGAACTC) (Choi et al., 2014). Amplifications were performed in 25 µl reactions with Qiagen HotStar Taq master mix (Qiagen Inc, Valencia, California), 1µl of each 5µM primer, and 1µl of template. Reactions were performed on ABI Veriti thermocyclers (Applied Biosytems, Carlsbad, California) under the following thermal profile: 95°C for 5 min, then 35 cycles of 94°C for 30 s, 54°C for 40 s, 72°C for 1 min, followed by one cycle of 72°C for 10 min and 4°C hold.

Products from the first stage amplification were added to a second PCR based on qualitatively determine concentrations. Primers for the second PCR were designed based on the Illumina Nextera PCR primers as follows: Forward - AATGATACGGCGACCACCGAGATCTACAC[i5index]TCGTCGGCAGCGTC and Reverse - CAAGCAGAAGACGGCATACGAGAT[i7index]GTCTCGTGGGCTCGG. The second stage amplification was run the same as the first stage except for 10 cycles.

Amplification products were visualized with eGels (Life Technologies, Grand Island, New York). Products were then pooled equimolar and each pool was size selected in two rounds

using SPRIselect Reagent (BeckmanCoulter, Indianapolis, Indiana) in a 0.75 ratio for both rounds. Size selected pools were then quantified using the Qubit 4 Fluorometer (Life Technologies) and loaded on an Illumina MiSeq (Illumina, Inc. San Diego, California) 2x300 flow cell at 10pM.

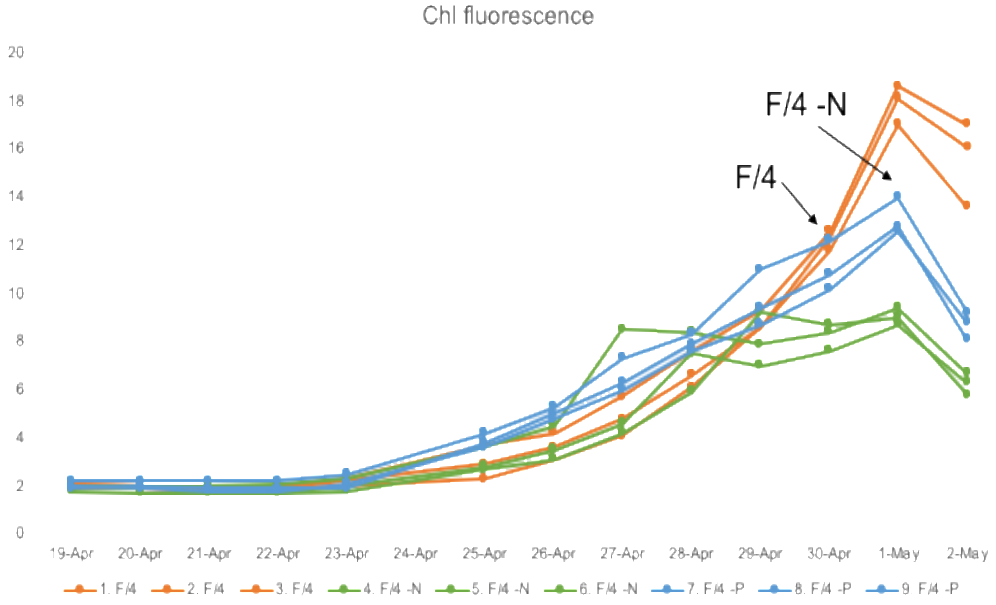
rpoC1 Clone Libraries

We amplified the *rpoC1* region using 25 ul reactions with 13.5 ul Qiagen GoTaq Hot Start Green Master Mix (Promega, Madison, WI, USA) 1 ul DNA from the extractions described above, and 1 ul of each 5 µM primer (SAN157F: 5'-YTNAARCCNGARATGGAYGG-3'; SAC-1039R:5'-CYTGYTTNCCYTCDATDATRTC-3') (Tai & Palenik, 2009). The reactions were denatured at 94°C for 2 min, then went through 30 cycles of 94°C for 30 seconds, 55°C for 1 min, 72°C for 1 min, with a final elongation step of 72°C for 10 min. After confirming presence of a band, we transformed the PCR product into TOP10 chemically competent *Escherichia coli* cells (Invitrogen, Carlsbad, CA, USA) via the pCR2.1-TOPO plasmid vector (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's protocols. We spread LB plates, incubated overnight at 37 °C, and selected blue colonies to grow overnight in LB broth containing 50 ug/ml Kanamycin. Plasmid DNA was extracted using the PureLink Quick Plasmid DNA Miniprep Kit (Thermo Fisher Scientific, Waltham, MA, USA) and sent to Eton Biosciences (San Diego, CA, USA) for sequencing using the M13F and M13R primers. Sequences were assembled using CLC Workbench v 8 (https://www.qiagenbioinformatics.com/).

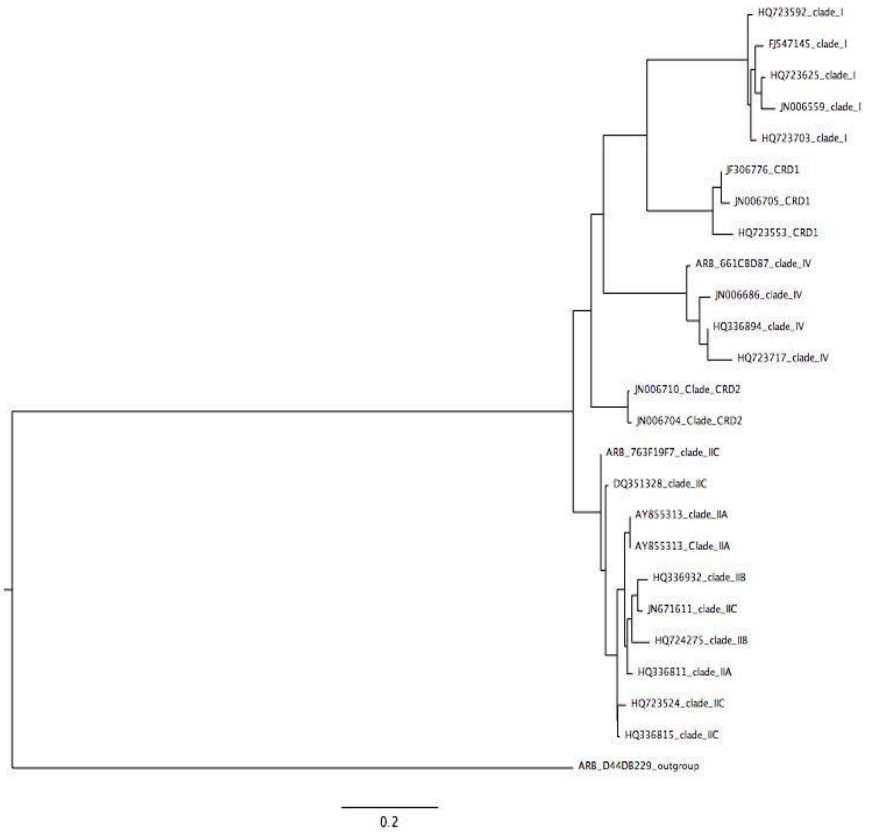
Supplemental table S4. Dates sequenced and taxonomic composition of rpoC1 clone libraries.

Date	Description	Num. Sequences	Num. Clade I	Num. Clade IV	Num. other
7 July 2011	Peak of 2011 bloom	28	20	8	0
21 July 2011	Decline of 2011 bloom	24	8	16	3
26 July 2012	Prior to 2012 bloom	16	8	8	0
2 August 2012	Peak of 2012 bloom	13	5	6	2

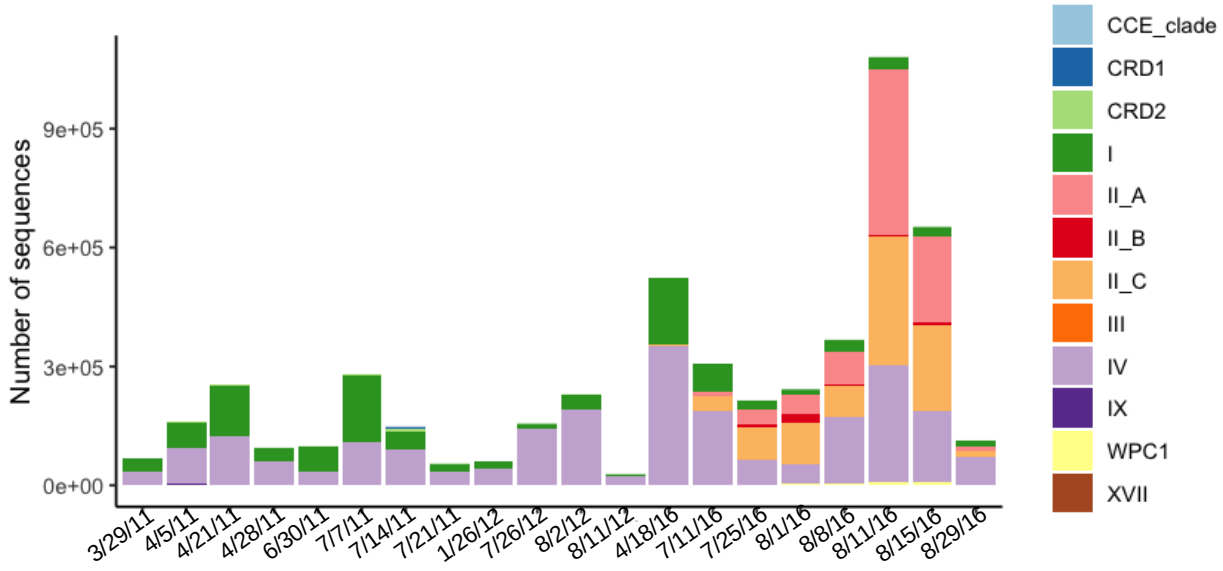
Sample	CC9311 proportion / cell #	CC9902 proportion / cell #	CC9305 proportion / cell #	CC9701 proportion / cell #	CC9605 proportion / cell #	Total cells in Mock community
MC 1	99%; 99,000,000	1%; 1,000				100,000,000
MC 2	99%; 99,000,000		1%; 1,000			100,000,000
MC 3	90%; 90,000,000	10%; 10,000,000				100,000,000
MC 4	25%; 7,500,000	25%; 7,500,000		25%; 7,500,000	25%; 7,500,000	30,000,000
MC 5	33.3%; 33,333,333	33.3%; 33,333,333	33.3%; 33,333,333			100,000,000



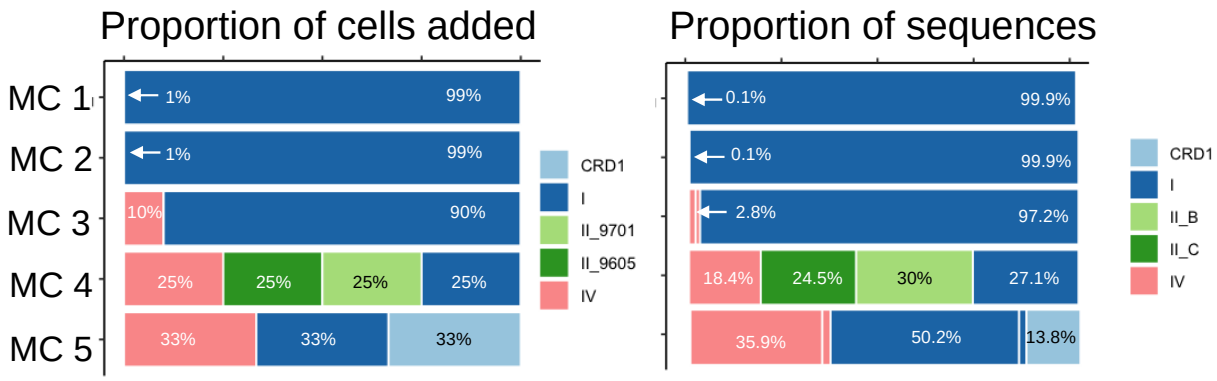
Supplemental Figure S5. (Top) Cell composition of five mock communities that were created and sequenced. (Bottom) Time points at which CC9902 cultures were harvested for DNA extraction.



Supplemental Figure S6. Maximum likelihood tree of sequences from the Choi et al. (2014) database to which abundant ASVs had highest percent identity. Names designate accession number of sequence along with the classification of the ASV it matched with according to the tree in Fig. 1. All clade II sequences cluster in this tree despite forming three clusters (II-A, II-B, II-C) in the ML tree made with only the amplicon region.



Supplemental Figure S7. Synechococcus community composition normalized to Synechococcus cells/ml.



Supplemental Figure S8. Comparison of known proportions of cells from each clade added to mock communities and resulting proportions of sequence reads from those clades. Clade I (CC9311) is consistently over-represented in sequence results resulting in slight under-representation of clade IV (CC9902) or clade CRD1.

CHAPTER 3:

Diversity and putative interactions of parasitic alveolates belonging to Syndiniales

3.1 ABSTRACT

The dinoflagellate lineage Syndiniales currently consists entirely of parasitic species that fall into five well-supported clades. Environmental sequencing studies worldwide have found an abundance of Syndiniales in a variety of marine ecosystems, but very little is known about the majority of Syndiniales species including two entire clades which have only been observed in sequence data. Syndiniales are known to have a wide range of hosts, but only a few dozen interactions have been confirmed through observation of actual infections. Here we describe the diversity of Syndiniales found at the Scripps Institutional Oceanography pier over the course of a year based on 18S sequencing. We find Syndiniales to be the most species (ASV)-rich taxonomic group and for its members to be present and abundant throughout the year. We use several analytical techniques to identify potential parasite-host interactions which we were then able to visualize over time. Using mock communities and size fractionation of seawater, we suggest that the majority of Syndiniales sequences that are found in environmental studies belong to the free-living dinospore stage rather than representing active infections.

3.2 INTRODUCTION

Parasitism is a species-species interaction where one organism both causes harm to, and is in some way nutritionally dependent on, a host that serves as its habitat (Anderson & May, 1978). It is vastly understudied among the millions of interactions occurring in marine microbial ecosystems. These interactions are challenging to observe as they often occur on microscopic visual and temporal scales. However, large-scale environmental sequencing projects have provided new methods of identifying interacting species. These are important to disentangle in any ecosystem because they inform our understanding of numerous factors, including: transfer of

carbon and energy, predicting vulnerability or robustness of given species, and understanding what confers fitness and how species adapt to differing biotic or abiotic conditions. It is necessary to expand our understanding of parasitism in marine ecosystems given its ubiquity as a trophic strategy and broad role in mortality and carbon cycling.

Current models of trophic webs include the 'microbial loop' – the concept of particulate organic matter getting released back into the water rather than continuing to the next trophic level due to factors like sloppy grazing and viral lysis – and the subsequent remineralization of that POM by heterotrophic bacteria. Because of this key discovery by Azam et al (1983), calculating the rates of grazing and viral lysis have become standard practices in understanding trophic efficiency. However, parasites also release particulate matter upon infection and lysis of their hosts. If parasitism constitutes a significant component of trophic interactions, then moving forward it must be incorporated as a major component in these processes as well.

There are multiple functional types of parasitism, some of which are considered a variation of predator-prey interactions. Specifically, parasitoids are a class of parasites that obligately kill their hosts in order to propagate and can thus in some ways be considered analogous to predators (Kuris, 1974). Castrators are parasites that disrupt and utilize the reproductive machinery of their hosts. As a source of mortality for phytoplankton and zooplankton, parasitism plays a role similar to predation, but is also responsible for putting organic matter back into the available supply – making it a very important process to understand in marine systems.

The parasitic group Syndiniales has been increasingly recognized due to its abundance and ubiquity in sequencing studies at numerous sites globally (Coats & Park, 2002; Guillou et al., 2008a; Park et al., 2013; Skovgaard et al., 2005, 2008; de Vargas et al., 2015). This is

believed to be a monophyletic and entirely parasitic lineage with several clades consisting entirely of uncultured species (Guillou et al., 2008b; Moon-van der Staay et al., 2001). These marine parasites have a wide variety of hosts, with known hosts including other protist parasites, ciliates, radiolarians, tintinnids, dinoflagellates (including harmful bloom formers), fish eggs, copepods, and crabs (Coats, 1999; Guillou et al., 2008b; Jung et al., 2016; Skovgaard et al., 2005; Stentiford & Shields, 2005). The Syndiniales comprise at least five major marine alveolate (MALV) groups, referred to interchangeably as MALV I-V, Syndiniales groups I-V (as in this paper's text, or (in the PR2 database, and thus in this paper's figures) Dino-Group I-V (Chambouvet et al., 2011; Guillou et al., 2008b; Guillou et al., 2013). Syndiniales group II is currently recognized as the most diverse clade and contains some of the most studied species, including members of *Amoebophrya.* Most of these have been identified as parasites of different dinoflagellates. Group I includes *Ichthyodinium*, a parasite of fish eggs, as well as members that infect ciliates, while group IV contains members that infect other metazoans (such as crabs and copepods), including *Hematodinium, and Syndinium.* Groups III and V currently consist entirely of environmental sequences, but form well-supported clades (Guillou et al., 2008b). Both functional types of parasitism are present within the Syndiniales; *Amoebophrya sp.* and other species that infect unicellular hosts are parasitoids, while some that infect metazoan hosts such as copepods act as castrators (Shields, 1994).

Some Syndiniales species have been studied more closely in lab-based experiments, demonstrating a broad range of hosts but a fairly similar lifestyle. An example of an *Amoebophrya* life cycle in species that infect dinoflagellates has three stages: dinospore, trophont, and vermiform (Coats & Park, 2002). The dinospore is an infective dispersal stage that can penetrate the pellicle of its host cell. Once it has infected the host cytoplasm or nucleus

(both of which have been observed but appear to be species-specific), it enters the trophont stage and replicates its own nucleus, forming a large multinucleate cell ("beehive"). Finally, it evacuates the host by rupturing its membrane and releases short-lived vermiforms, which must complete cytokineseis in order to become biflagellated dinospores, which are much smaller than most known Syndiniales hosts (Coats & Bockstahler, 1994).

There is evidence that parasitism may play a role in the decline of dinoflagellate blooms. This has even been suggested as a means of controlling harmful algal blooms since many toxic dinoflagellates have been identified as Syndiniales hosts (Taylor, 1968). In the Penzé estuary in France, Chambouvet et al. (2008) found successive blooms and declines of four different dinoflagellate species, *Heterocapsa rotundata*, *Scrippsiella troichoidea*, *Alexandrium minutum*, and *Heterocapsa triquetra* over the course of weeks. They were able to visually track Syndiniales infections using fluorescent probes specific to MALV II (Syndiniales group II), as well as sequencing at different time points, which revealed prevalence of different Syndiniales species during or shortly lagging each of the four dinoflagellate blooms. This type of succession pattern was detected consistently in multiple years. Montagnes et al. (2008) comparatively modeled the dynamics of harmful algal blooms with just microzooplankton grazers present versus grazers and parasites, and found that the second scenario better reflected actual bloom dynamics observed in the Penzé estuary. The contribution of parasitic infection to the decline of a bloom is typically represented as host mortality, the percentage of hosts killed per day (Velo-Suárez et al., 2013), with counts generally conducted via epifluorescence microscopy (Coats et al., 2002; Salomon et al., 2003). Mortality is estimated by dividing the number of observed infections by the infection time (D. Wayne Coats & Bockstahler, 1994; Mazzillo et al., 2011),

and comparing this to the actual rate of decline of the host population can quantify the contribution of parasitism.

Known parasites comprised more than half of both the richness and abundance of sequences within the piconanoplankton collected in the global TARA expedition, which sampled at 68 stations across the global oceans (de Vargas et al., 2015). At our own site, the Scripps Institution of Oceanography Pier, Syndiniales sequences comprised as much as 11% of sequences at a given sampling point (Nagarkar et al., 2018). In the Californa Current Ecosystem, Mazillo et al. (2011) found that *Akashiwo sanguinea* blooms were associated with low *Amoebophrya* prevalence, while non-red tide periods contained a higher percentage of infected *Akashiwo* cells. One of the years with an especially long red tide, no *Amoebophrya* were detected, which indicates an important role for these parasites in maintaining *Akashiwo sanguinea* populations at a low level. It is possible a similar phenomenon occurs at the SIO pier; 29 Dec 2011, the date of the largest *Akashiwo sanguinea* bloom, had one of the lowest number of *Amoebophrya* and Syndiniales sequences. This bloom could have been a result of *Akashiwo sanguinea* growth becoming greater than its mortality rate due to some factor leading to *Amoebophrya* decline. There is evidence that ciliates graze on *Amoebophrya* dinospores, reducing their infection rates on dinoflagellates such as *Akashiwo sanguinea* (Johansson & Coats, 2002), so this is one possible factor.

However, there are many unknowns when it comes to using sequence data to study this abundant group. For example, it is unclear whether their abundance in sequence data truly translates to environmental abundances. This is complicated by the issues of copy number, and the fact that their different life stages include numerous nuclei when inside a host, as well as dinospores which may not go on to infect anything (and have been found in some species to

survive for approximately two weeks in their free-living state). Each of these life stages consists of different cell sizes and survives for a different amount of time. For one species of *Amoebophrya* that was observed to infect *Gymnodinium sanguineum*, dinospores were less than 10 μm in size and survived independently on the order of days; trophonts were multinuclear and reached up to 20 μm, with infections that also lasted on the order of days; vermiforms appear as long connected cell chains, with a length possibly in the hundreds of microns, but are released as dinospores within minutes (Coats & Bockstahler, 1994; Coats & Park, 2002). Ultimately, the ratio of life stages represented by sequence data is unclear; perhaps a single infected cell containing a trophont can yield many reads of the 18S sequence, or perhaps the majority of these come from dinospores but cannot provide information about how many actually cause infections in new hosts. Dinospores can be observed in their free-living state (Siano et al., 2011) and thus could potentially be quantified.

One approach to this question is through co-occurrence analyses, though for the most part these have been geographical. For example, the TARA Oceans expedition, using co-occurrence matrices made from OTU abundance data at numerous sites, found that parasitism was the most abundant type of taxon- taxon interaction (Lima-mendez et al., 2015). A majority of these interactions involved the Syndiniales, especially clades I and II. Other studies have encountered unexpected abundances and diversity of Syndiniales in a variety of environments, including the Antarctic (Cleary & Durbin, 2016), the Baltic Sea (Majaneva et al., 2012), and Korean coastal waters (Kim et al., 2008). However, we know marine communities to be dynamic over time especially at coastal sites, and geographical co-occurrence should be supplemented with time-series studies at a given site which may be able to validate or reveal new interactions. Syndiniales sequence abundance and phylogenetic composition are temporally variable at the

pier, and seasonality in the clade distribution of Syndiniales has been observed at other sites as well (Chambouvet et al., 2008; Chambouvet et al., 2011; Cleary & Durbin, 2016). Most approaches identify interactions by finding correlations between species; however, in reality not all interacting pairs will be correlated and may exhibit a time lag or anti-correlation. More recently, methods for identifying causal interactions using simplex projection and convergent cross mapping (CCM) have been proposed to address this (Sugihara et al., 2012; Ye et al., 2015). A major challenge of any of these methods is the fact that environmental sequence data is ultimately compositional, and knowing true abundances would be most ideal for understanding temporal changes in communities. There have been several proposed workarounds to this, including: spike-ins, analyses optimized for compositional data, normalizing using other forms of data collection, etc. While no method is perfect, we have attempted to use multiple approaches to gain an understanding of what is happening at our site.

In this study we use high-frequency environmental sequencing at a single site to describe the dynamics of the eukaryotic community as a whole and attempt to identify individual parasite-host interactions that might be ecologically relevant. We expect that Syndiniales diversity will vary throughout the year based on its host populations and that parasite-host pairs might have high co-occurrence. We also predict that some parasite-host pairs will show signatures of causal interactions.

3.3 METHODS

3.3.1 Sample collection and DNA extraction

We sequenced 87 seawater samples collected from the Scripps pier along with several additional replicates and mock communities. For sampling, surface seawater was collected by

bucket at the end of the Scripps pier (32º87'N, 117º26'W) on every Monday and Thursday (with one exception, 4 July) in 2016 and filtered onto a 47 mm 0.2 µm Supor filter (Pall, New York, USA) unless otherwise indicated. Triplicate filters were wrapped in aluminum foil and stored at -80ºC until the time of DNA extraction. DNA extractions were conducted in sets of two to four filters. Filters were cut into small pieces on a clean surface and placed in 2 ml tubes with 560 µl TE (50 mM tris, 20 mM EDTA) and 80 µl of 100 mg/ml lysozyme. *S. pombe* DNA was spiked in a this time for some of the samples (explained in further detail in supplementary). Samples were incubated at 37ºC for half an hour. We then added 10 mg/ml Proteinase K and SDS and incubated at 55ºC for 2.5 h. Finally we added 10 µl 10 mg/ml RNAse A and incubated at 37 ºC for another half hour. DNA was extracted in an equal volume of phenol:chloroform:isoamyl alcohol twice and then once more with chloroform:isoamyl alcohol. Finally we eluted DNA using the Qiagen DNEasy Blood and Tissue kit (Qiagen, Germany) according to the manufacturer's instructions. DNA was stored at -20ºC until further use.

3.3.2 Library preparation

For approximately half the samples, triplicate 25 µl PCRs were performed on each sample using the Euk_1391F (5'-GTACACACCGCCCGTC-3') and EukBr (5'-TGATCCTTCTGCAGGTTCACCTAC-3') primers with single-index barcodes on the forward primers. Reactions used 1 µl of each primer and 10 µl of GoTaq HotStart Master Mix (Promega, Wisconsin, USA). The triplicate reactions were combined and DNA concentration was measured using a Qubit fluorometer. Unincorporated nucleotides were removed using the ExoSAP-IT cleanup kit (Thermo Fisher Scientific, Massachusetts, USA) according to the manufacturer's instructions. All sample DNA was pooled into a single tube with approximately 220 ng per sample, and the pooled sample was cleaned with Agencort AMPure beads (Beckman

Coulter, Indiana, USA) according to the manufacturer's protocol. This sample was sent to the IGM for paired-end 150 sequencing on the Illumina MiSeq.

Other samples were sent to RTL Genomics (Lubbock, TX, USA) for sequencing and processed according to their protocols (S1).

Schizosaccharomyces pombe DNA was added to many of the samples before or after the DNA extraction step, but was not used in the final analysis. A brief description of the methods and rationale for not using the sequence results is provided in S2.

3.3.3 Mock communities and replicates

We sequenced several mock communities along with the seawater samples. Mock communities were created by counting cells from different eukaryotic lab cultures and pipetting known proportions of cells onto a 0.2 µm filter. Subsequently filters were treated as though they were an environmental sample (as described above). Additionally, we sequenced replicates of certain samples (including the same sample sequenced at RTL genomics and using our own library prep). Results from these are available in S3 and S4.

3.3.4 Normalization of sequences using picoeukaryote flow cytometry counts

Flow cytometry of samples was conducted as described in Nagarkar et al. (2018). Counts from a gated region known to contain photosynthetic picoeukaryotes, specifically members of Mamiellales (*Micromonas, Ostreococcus, Bathycoccus*) were used to adjust the sum of all Mamiellales sequences from the 18S data. Then this adjustment factor was applied to all other ASVs within the sample.

3.3.5 Data analysis

Sequences were assigned to ASVs (amplicon sequence variants) using deblur (Amir et al., 2017), with default parameters and a trim size of 155. Resulting ASVs were classified using

BLCA (Gao et al., 2017) against the PR2 database (Guillou et al., 2013). Most subsequent analyses were conducted in R using the phyloseq (Mcmurdie & Holmes, 2013) and rEDM (Ye et al., 2018). Correlation analyses were conducted using SparCC (Friedman & Alm, 2012). *S. pombe* sequences were removed prior to all analyses and unclassified or metazoan sequences were removed for some subsequent analyses. For both of the following analyses we have reported interacting pairs with a cutoff of r = 0.6 or greater unless otherwise specified.

Prior to running SparCC, ASVs with fewer than 100 sequences were eliminated, along with ASVs with greater than 60 zeroes. This was a slightly more permissive cutoff than that used for convergent cross-mapping (CCM, described next) because SparCC is intended for sparse datasets. After these cutoffs a total of 1043 ASVs were run against one another to find significant correlations. Subsequently, 100 simulated datasets were created and SparCC was run on the bootstrapped datasets as well to obtain p-values for significant correlations.

ASVs were chosen for CCM analysis based on the following standards: (1) Unclassified and metazoan ASVs were eliminated. (2) Any ASV with >50 zeroes was eliminated. (3) Any ASV with <10000 total reads among all time points (after picoeukaryote-adjustment). These cutoffs were set for optimizing our interpretation of results; it has been previously shown that nonlinearity is harder to detect in time series with more zero values (Giron-Nava et al., 2017). For each ASVs used, simplex projection with leave-one-out cross-validation was used to identify the best embedding dimension based its on forecasting ability. This embedding dimension was used in convergent cross mapping to test how well time-lags one series could predict another, with an exclusion radius (an indicator of causality). In each case the Pearson correlation coefficient between predicted and actual values, rho, was tested for significance against 100 surrogate datasets. In addition to running CCM on individual ASV-ASV interactions, we

aggregated ASVs belonging to the same taxonomic groups and ran these against one another as well as against several environmental parameters (S5).

3.4 RESULTS

3.4.1 The eukaryotic community at the Scripps pier over time

From 87 samples containing 5,181,108 sequences we found there were 5632 ASVs at the pier in 2016. As expected, there were fluctuations in the community composition at both broad and fine taxonomic levels (Fig. 3.1). Several phenomena were immediately observable in these sequence data including a dinoflagellate bloom in early July followed by raphidophyte bloom starting in July that was corroborated in observational data (Fig. 3.1).

Only six ASVs were present in every single sample: one cryptophyte (ASV 4, *Teleaulax sp.*), two dinoflagellates (ASV 9, *Heterocapsa sp.*, and ASV 145, *Biecheleria sp.*), and three chlorophytes, (ASV 25, *Bathycoccus sp.*, ASV 44, *Micromonas pusilla*, and ASV 34, *Pyramimonas sp.*). Classifications provided in parentheses were the top BLAST hit for each ASV sequence. The vast majority of ASVs were present at fewer than half the time points, with certain groups like apicomplexans and radiolarians comprised mostly of ASVs that were detected on fewer than 10 time points (Fig. 3.2). Among the 50 ASVs that were present for greater than 90% of time points, four belonged to the Syndiniales—ASVs 14, 76, 233, and 707. ASV 14 was classified in Syndiniales group III and the others were in group I.

An ordination analysis of all time points reveals there to be a seasonal structuring of the community (Fig. 3.3), which might help explain the lack of ASVs present at all times.

3.4.2 Use of picoeukaryote counts to quantify sequence data

The compositional nature of sequence data makes it possible for an ASV to decrease in relative abundance when it has actually increased in absolute abundance – for example, when another member of the community blooms, it decreases the relative abundances of everything else. To mitigate this, we adjusted read counts based on abundances of cells counted within a known picoeukaryote-containing gate using flow cytometry. This was done by normalizing all read counts falling within *Mamiellales* to the picoeukaryote flow cytometric counts (Fig. 3.4A, B) and applying this correction factor to all sequences (Fig. 3.4C). Additional discussion of this quantification method can be found in S2. We hope that this method allows us to better track "absolute" abundances of organisms relative to *one another* – as in, determine whether something is increasing or decreasing between two time points. However, previous knowledge of copy number variation, and results of our own mock communities (S4) make it clear that we still cannot determine relative numbers of cells to one another based on numbers of sequences; for example, in the case of our mock community, the dinoflagellates *Lingulodinium* and *Heterocapsa* were greatly overexpressed relative to all other members of the community.

3.4.3 Syndiniales sequences are highly diverse and abundant at the pier

There were 997 total Syndiniales ASVs, making it the most diverse taxonomic group (aside from "unclassified"). There were hundreds more Syndiniales ASVs than any other ubiquitous taxa, including the stramenopiles, dinoflagellates, ochrophytes, metazoans, chlorophytes, cercozoans, and haptophytes (Fig. 3.5A). Syndiniales sequences were not typically the most abundant in samples, ranging from 2 to 20% of total sequences during a given time point; the greatest proportion of total sequences was usually from other dinoflagellate taxa (Fig. 3.5B). Syndiniales clades I and II were the most abundant, with clades III, IV, and V present at lower levels – though ASV 14, of clade III, was present nearly all the time. The

Syndiniales proportion of reads was greatest in the summer dates (Fig. 3.6A), but individual ASVs had distinct times of year when they were present, with very few present all throughout the year (Fig. 3.6B-E).

3.4.4 Identification of putative Syndiniales-host interactions: Co-occurrence and correlation

Given that currently characterized members of Syndiniales spend a significant amount of their life cycle within their host, it is likely that host-parasite pairs would co-occur. The Syndiniales-*Karlodinium* pair that has been maintained in culture (Bai et al., 2007, obtained from *T. Bachvaroff*) and was used in the mock communities did appear, and co-occur, in our pier samples, but both sequences were only present abundantly on a single date (9 June 2016), so their relationship cannot be tracked temporally in this dataset.

We conducted several different time series analyses to identify interactions among Syndiniales ASVs and other ASVs, which would be potential hosts. Using SparCC (Friedman & Alm, 2012), we found 664 significant ($p < 0.01$) correlations with $r > 0.6$, and 135 of these involved a Syndiniales ASV. There were correlations among Syndiniales ASVs and a wide variety of other ASVs, including ciliates (10), raphidophytes (8), diatoms (20), dinoflagellates (21), chlorophytes (21), cercozoans (10), and other members of the Syndiniales (11). Four examples of highly correlated Syndiniales-host pairs are provided in Fig. 3.6B-E. It is noteworthy that the adjusted counts of Syndiniales sequences compared to putative host sequences differ amongst each of these. For example, there were typically more Syndiniales sequences than prospective host sequences in Fig. 3.6B (ASV 586 and 944), but this was not always the case. Importantly, the SparCC correlations were obtained based on proportions but in Fig. 3.6 the ASV patterns are shown based on "total" (picoeukaryote-adjusted) abundances. A full list of highly correlated pairs is provided in S6.

3.4.5 Identification of putative causal interactions between Syndiniales and other species

We used empirical dynamical modeling to identify additional causal interactions that are not detected by traditional correlation approaches. Through convergent cross-mapping we found 85 significantly causal interactions whose signals were robust to two different metrics: (1) CCM using a time series with weekly intervals for all of 2016 (2) CCM using a second, shorter time series, which was also weekly but on a different day of the week (Fig. 3.7A). Members of Syndiniales were causal to other ASVs as well as causally driven by other ASVs. Additionally, 19 of these interactions were also significantly positively correlated (with $r > 0.6$) in SparCC, three of which included a member of Syndiniales (Fig. 3.7B). Of these, two involved the same Syndiniales ASV causally interacting with another (a dinoflagellate and a raphidophyte). While it is not expected that causally interacting pairs would necessarily also have strong correlation, we suggest interacting pairs that are robust to both of these tests might be particularly promising for further investigation.

3.4.6 Coupling of microscopic observations and amplicon data reveals other putative pairs

In March 2018, there was a large diatom and *Ceratium* dinoflagellate bloom at the SIO pier with readily observable *Ceratium* infections (Fig. 3.8A). We sequenced this date and found a *Ceratium* ASV (ASV 330) and Syndiniales clade II ASV (ASV 947) that were highly abundant. The Syndiniales clade II ASV had a 100% match on BLAST with *Ameobophrya ex. Ceratium tripos* (Accession no. AY208892). When tracked in the 2016 time series (Fig. 3.8B), these ASVs did co-occur and indeed were both particularly abundant on the same date, but were not identified as an interacting pair given the SparCC correlation threshold of $r = 0.6$ used in the previous methods (they had a significant but low SparCC correlation of $r=0.26$).

3.4.7 Representation of Syndiniales life stages in sequence data

Based on sequencing mock communities containing only *Karlodinium* and *Amoebophrya* we see that host sequences are over-represented (over 75% of total sequences) even when the sample contains an equal number of *Karlodinium* cells and dinospore cells (S 3.7). Sequencing environmental samples obtained through a 0.2 µm filter captures both dinospores and active infections. In order to properly track the progression of infections, it is necessary to size-fractionate seawater. To examine whether this would be a viable method, we conducted a pilot study wherein we compared samples filtered on a 10 µm filter, on a 0.2 µm filter, and on a 0.2 µm filter after pre-filtration on a 10 µm filter (Fig. 3.8A). We found that 0.2 µm samples were enriched in Syndiniales sequences as a proportion of total dinoflagellate sequences, with the 10 µm pre-filtered sample even more enriched than the other 0.2 µm sample (which makes sense given that there would be fewer cells captured in this sample, resulting in greater proportional representation of whatever is captured). This likely indicates that active infections have been excluded from samples marked "3" in the figure. Comparison of 10 µm and 0.2 µm filters from the same seawater sample indicates that dinospores accounted for more of the sequence reads than infections, as Syndiniales sequences were close to 25% of all dinoflagellate sequences on a 0.2 µm filter, but dropped to being less than 10% on the 10 µm filter (Fig. 3.9B). This applied not only to Syndiniales as a group but also to nearly every ASV. These preliminary data indicate that the majority of Syndiniales sequences on 0.2 µm filters might represent the free-living dinospore life stage. However, an additional caveat is that there were a number of dinoflagellate ASVs that passed through the 10 µm filter (as in, were found in sample D but not sample C as designated in Fig. 3.9B). These included ASVs classified within *Gymnodinium*, *Heterocapsa*, and *Prorocentrum*. So some potential hosts may not be able to be separated using this method and other combinations of filter sizes may be more useful.

While Syndiniales group III and IV were present at low abundances overall, their sequences were *only* found in the smaller size fraction (none in the >10 µm size fraction). In the case of clade IV, which includes *Hematodinium* and *Syndinium*, it makes sense that we may not have captured any eggs or larvae of crustaceans which are the known hosts of these species. In the case of clade III, this group is yet uncharacterized and its hosts are not known. However this result suggests that either (a) these samples occurred after an infection cycle where no infected hosts were present but some dinospores still persisted, (b) these also infect larger organisms (metazoan eggs or larvae) that would be less likely to be captured in a given 500 ml seawater sample, or (c) these infect hosts smaller than 10 µm in size.

Finally, though we only have two time points, one week apart, where both size fractions were sequenced, we are able to observe the progression of infections in some Syndiniales ASVs. Three examples are provided in Fig. 3.9C. ASV 1035 was only detected in the 0.2 µm filter during the first week but detected on the 10 µm filter in the second week, indicating that dinospores were present initially and may have begun infecting hosts by the second week. A similar trend was seen with ASV 3941, with the proportion of infected hosts increasing by the second week. In the case of ASV 917, the mortality rate decreased by the second week.

3.5 DISCUSSION

We chose to explore the dynamics of Syndiniales populations at our site because of their previously suggested role in driving population dynamics of their hosts (Chambouvet et al., 2008; Coats et al., 1996; Coats & Bockstahler, 1994), particularly dinoflagellates, which are known to be abundant and have large blooms at our site (Nagarkar et al., 2018). If abundance of rDNA sequences is any indication (as our mock communities containing *Amoebophrya* suggest),

then it is clear that members of Syndiniales are a large component of marine microbial communities including our own. Yet, relatively little is known about the ecological strategies or even simply the hosts of most Syndiniales species. Members of Syndiniales were found to have the greatest number of species interactions within numerous ecosystems in a large-scale co-occurrence-based geographic study (Lima-mendez et al., 2015). Our ability to track and potentially enumerate Syndiniales populations at a single site can provide further insights into species-species interactions.

The SIO pier contained a large number and diversity of Syndiniales sequences. Many sequences appeared transiently for a few weeks or months but were absent the rest of the year. Members of Syndiniales did not ever constitute the largest proportion of sequences in a given sample, but were consistently abundant. Syndiniales was, however, the taxonomic group with greatest species (ASV) richness, with the majority of ASVs falling within clade II. While this remarkable diversity is unsurprising given the worldwide ubiquitous detection of Syndiniales sequences, our ability to classify these sequences is still very limited. We have reason to believe that the majority of Syndiniales sequences might represent free-living dinospores and would expect that their appearance at a given time is indicative of the presence of its host(s) as well. As they have not been documented to survive outside of their host for longer than a few weeks, their disappearance would indicate the absence of their host, possibly due to high mortality from infections. However, several members of Syndiniales (particularly *Amoebophrya*) are known to have multiple hosts (Coats & Park, 2002); thus we cannot assume near-absolute co-occurrence (even with a time lag) as we might if there were only single host specificity.

We chose to apply several different normalization methods and analysis tools to our sequence data. It has been previously found that less than half of edges are common to outputs

from different correlation methods (including LSA, SparCC, Pearson, and more) when using these to assess the same microbial community (Weiss et al., 2016). Additionally, the majority of microbial analysis tools are (1) based on identifying correlations and (2) applied to compositional data. Here we used SparCC, which is intended for sparse compositional data, but explored some of the significant interactions by looking at the ASV time series as "absolute counts" (by adjusting the sequences based on picoeukaryote abundances). Additionally, we explored not just correlation but also identified putative causal interactions by determining whether time-lagged embeddings of each time series could predict one another, a sign that the two might be causally linked (Sugihara et al., 2012).

A recently published database of observationally or experimentally-confirmed protist interactions obtained from an extensive literature review found 200 out of 2422 interactions involved parasitism by a member of Syndiniales (Bjorbækmo et al., 2019). Half of these were between the clade II genus *Amoeobphrya* and dinoflagellates, with a few more between *Amoebophrya* and ciliates or acantharians. The rest were between Syndiniales clades I or II and rhizarians, cercozoans, ciliates, chrysophytes, or dinoflagellates. Notably, no interactions have been specifically observed or characterized with protists and Syndiniales clades III, IV, or V (though clade IV members have known metazoan hosts). Among those characterized in the database, four Syndiniales species were observed to have more than one host (in three cases, multiple dinoflagellate hosts, and in one case multiple ciliate hosts). *Amoebophrya* spp. have almost exclusively been associated with dinoflagellates with the exception of a few interactions with ciliates and acantharians. Clade I and the remaining members of clade II were only identified to interact with rhizarians. In contrast to comprising one-tenth of the interactions in the PIDA database, Syndiniales parasites dominated the co-occurrence-based associations

identified by the TARA oceans study (Lima-mendez et al., 2015). In our co-occurrence analysis we found approximately one-sixth of the correlation-based interactions (that were significant with $r > 0.6$) included a member of Syndiniales. We found correlations between Syndiniales ASVs and dinoflagellates, ciliates, cercozoans, other syndinians, chlorophytes, radiolarians, prymnesiophytes, ochrophytes, katablepharids, and members of MAST.

We found causal interactions between Syndiniales ASVs and dinoflagellates, diatoms, chlorophytes, cercozoans, telonemians, ochrophytes, prymnesiophytes, picozoans, and members of MAST, but interestingly only a single ciliate ASV. Given the documented relationships between ciliates and members of Syndiniales, both of parasitism and of predation dinospores by ciliates (Johansson & Coats, 2002), we would expect causal interactions between these groups at our site. However, given the transience of many ASVs, some may not have been detected or their time series may have contained too many zeroes to be detected using CCM. Additionally, it is very possible that some declines in a host population are due to infection prevalence while others are driven by separate biotic or abiotic factors. – these would be overlooked by correlation analyses. There were four putative causal interactions between pairs of Syndiniales ASVs, which might be indicative of competition between different species as a driver of their population dynamics (as opposed to correlations which might indicate that those ASVs actually represent the same species).

While Syndiniales groups I and II had considerably more ASVs and sequences, there were 39 group III ASVs and some were present throughout the year. This group clearly has a significant and persistent presence at our site and yet almost nothing is known about it. The ubiquitously present Syndiniales clade III ASV, ASV 14, was correlated with ASV 25 (*Bathycoccus*), ASV 171 (Katablepharidophyta), ASV 40 (Mamiellales) and ASV 133

(Dictyophyceae). CCM revealed ASV 14 to have a putative causal interaction with another Syndiniales ASV (ASV 87, clade II). We found several other causally-driven interactions between members of Syndiniales group III and other ASVs (MAST-3, two members of Telonemia, and one Cercozoan) that also had significant correlations (ranging from 0.47 to 0.58) using SparCC (S6). To our knowledge no specific associations involving Syndiniales group III have been reported elsewhere thus far.

Our effort has yielded a number of likely interactions, which are now important candidates for further validation on environmental samples. We know some of the taxa which are parasite-host pairs based on our data, such as the dinoflagellate host in Fig. 3.6C, *Gymnodinium* with *Amoebophrya* , have been previously validated on the genus level (Coats et al., 2010) – but limitations in the length of our amplicon and the current sequence databases do not allow for confident classification at higher taxonomic resolution. We were able to look in our sequence dataset for two other known parasite-host pairs: one that has been maintained in culture between the dinoflagellate *Karlodinium* and *Amoebophrya*, and one pair that was clearly observed microscopically during a large dinoflagellate bloom (Fig. 3.9A). In both these cases, we did find convincing co-occurrence of the ASVs (for the latter pair, these were extrapolated from their abundance in the bloom sample), but neither pair was identified using the SparCC or CCM analysis techniques. This too is likely due to their transience; both were present in abundance for a very small number of time points. Given the clear seasonality of our site (Fig. 3.3), a multi-year time series might allow for more non-zero time points for a given host that only appears for a small part of the year.

Another crucial confounding factor in the use of sequence-based time series analyses to identify parasitic interactions is the lack of partitioning between parasite life stages. One way to

mitigate this problem and study individual time series would be size-fractionation of seawater prior to DNA extraction. From a preliminary comparison of two size fractions, we were able to separate the dinospore sequence signal from that of active infections. We found cases, like ASV 1035, where Syndiniales ASV sequences were not present on a 10 µm filter but were detected on a 0.2 µm filter. As we would expect only active infections (captured within a larger host cell) to be captured on a 10 micron filter, with the smaller dinospores passing through, this would indicate that only dinospores were present in that sample. Sequencing the same sample at these two different size fractions can provide information about what proportion of the sequences represent infections and what proportion represent dinospores. As demonstrated by our preliminary size-fractionation data (Fig. 3.8), looking at two size fractions over time could enable tracking infection prevalence and mortality rates – even if the host is unknown. The TARA oceans study reported size-fractionated sequence data and found a large majority of Syndiniales clade I and II sequences to be in the pico-nano (0.8 to 5 µm) fraction (de Vargas et al., 2015), so it is likely that most Syndiniales sequences represent dinospores rather than infections.

Members of the Syndiniales clearly play an active role in the microbial ecosystem at the SIO pier. In addition to the simplistic single parasite-single host interactions, it is known that some of these might be generalists in terms of their hosts. In a system with hosts vulnerable to multiple parasite species and parasites lethal to multiple host species, there might be a dilution of signal when analyzing time series. Additional dynamics, like competition between Syndiniales species for host resources, and even consumption of dinospores by protists, have not been properly accounted for in ecosystem-level analyses. We suggest that high-frequency and high-throughput environmental sequencing has the potential to help track dynamics of parasites and

hosts, but only with careful consideration of how to enumerate the cells and distinguish among their life stages.

3.6 FIGURES

Figure 3.1 Eukaryotic community at the SIO pier over the course of 2016. Metazoan sequences, unclassified sequences, and *S. pombe* sequences were removed, and replicate samples were pooled.

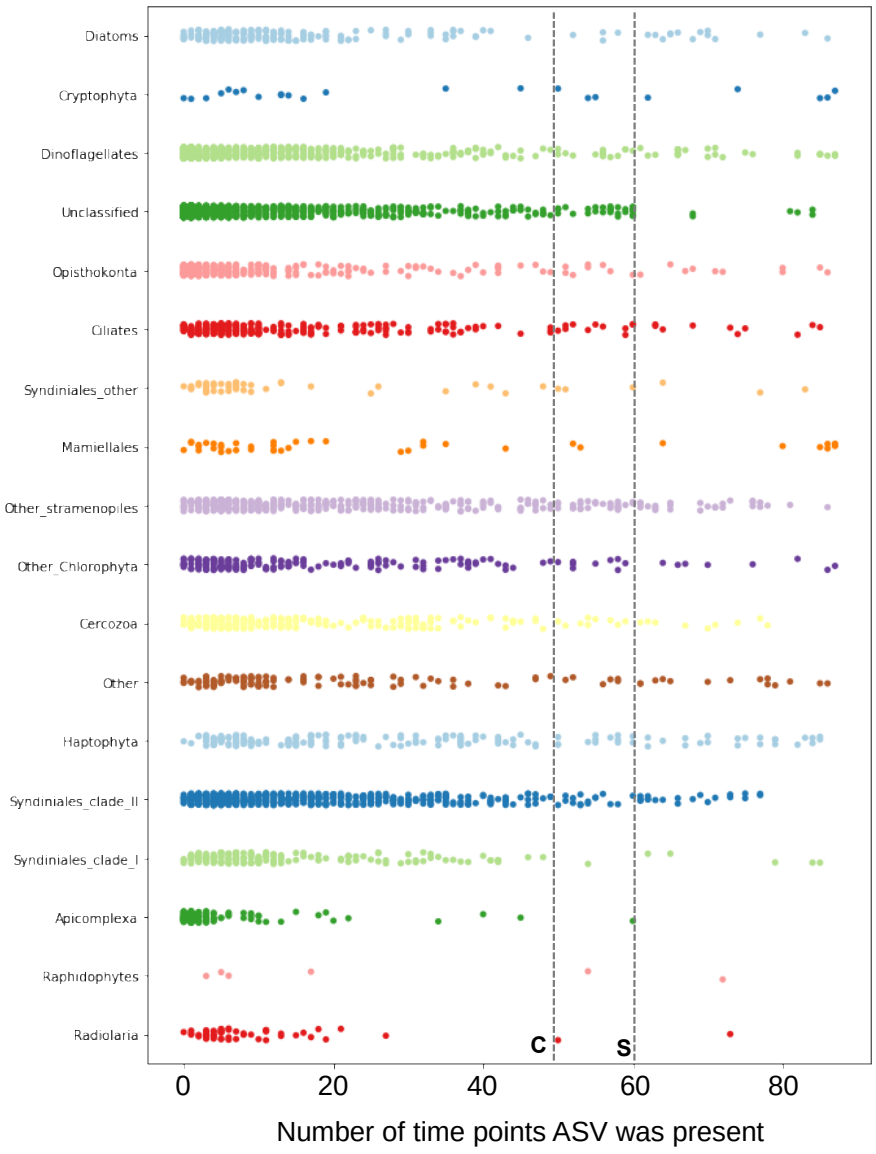


Figure 3.2 Number of time points at which ASVs were detected, grouped by taxonomic level. Each point represents an ASV (and points are jittered for clarity). Unclassified sequences were not removed and may include bacterial and archaeal sequences. "Other" refers to all other taxa not included in the groups. Dotted lines demarcate the cutoff for SparCC (S) and CCM (C). Only ASVs to the right of the lines were used for their respective analyses (though additional cutoffs were used as described in the methods).

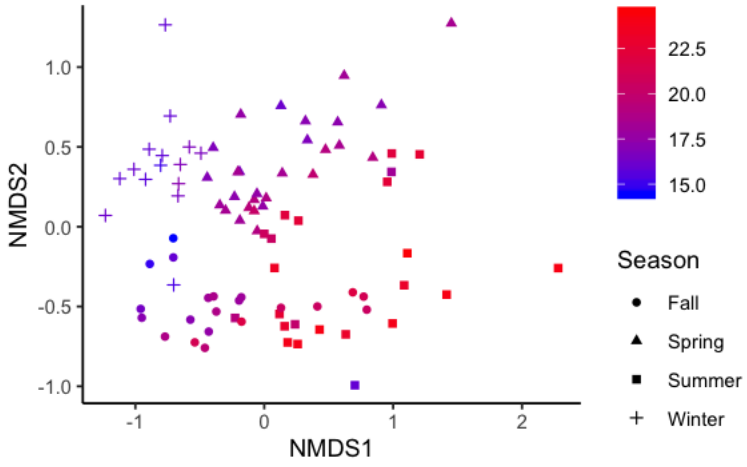


Figure 3.3 NMDS plot of all SIO pier samples, colored by temperature. Shapes represent the calendar season of the sample.

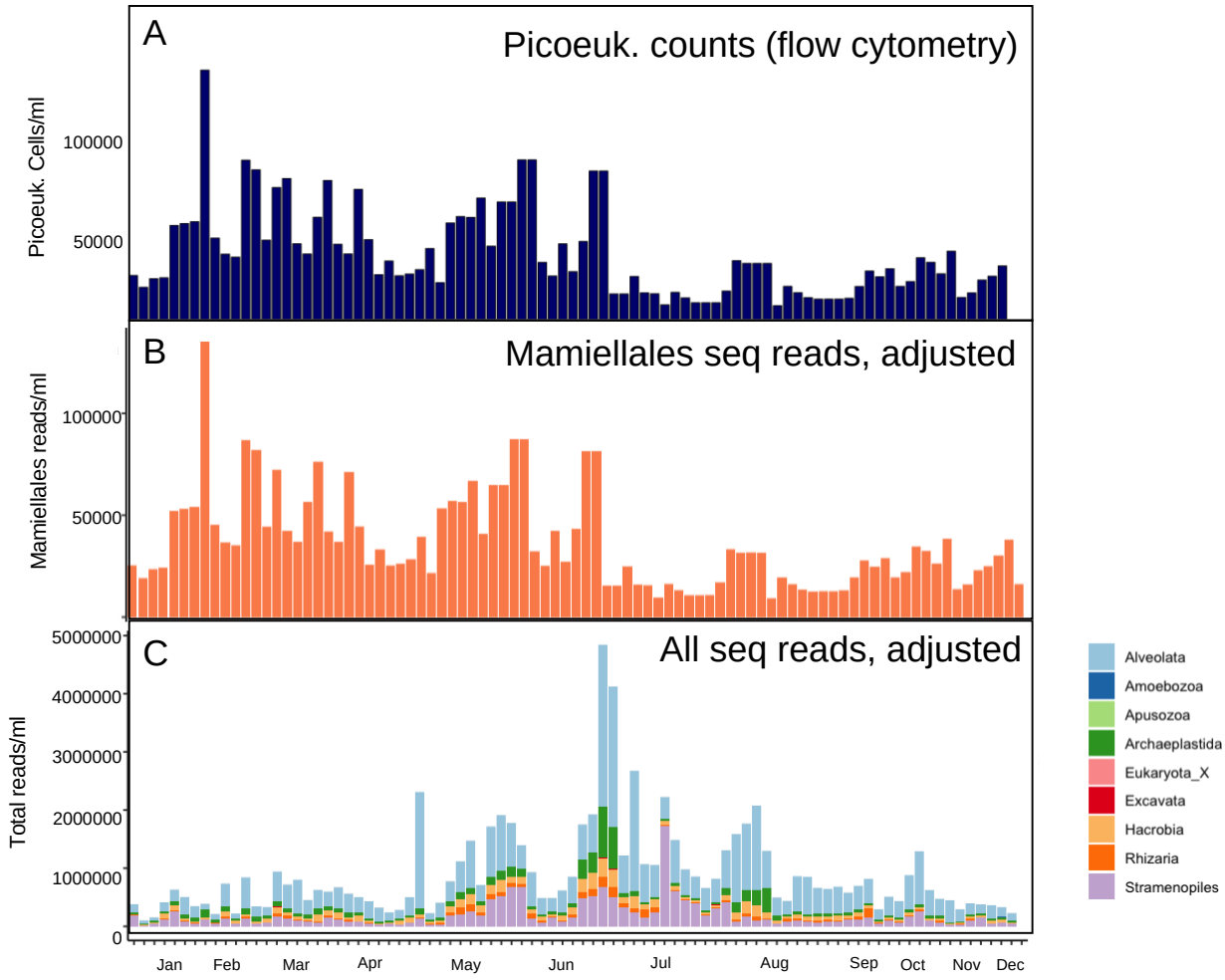


Figure 3.4 (A) Flow cytometric counts falling within the picoeukaryote gate (cells/ml). (B) Mamiellales sequence reads at each sample adjusted using the flow cytometry counts. (C) All sequences adjusted based on the correction factors obtained using (B). The identical bars in A and B represent replicate samples.

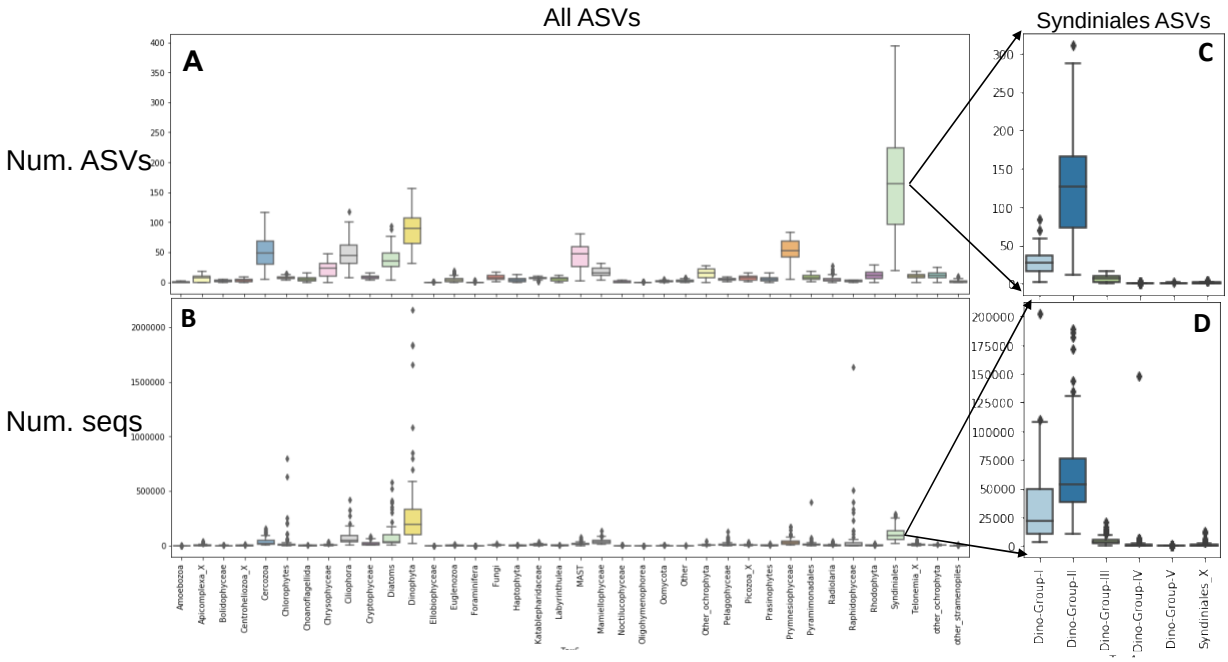


Figure 3.5 (A) Number of ASVs and (B) number of sequences within the different taxonomic groups amongst all pier samples. Number of sequences have been adjusted using picoeukaryote counts.

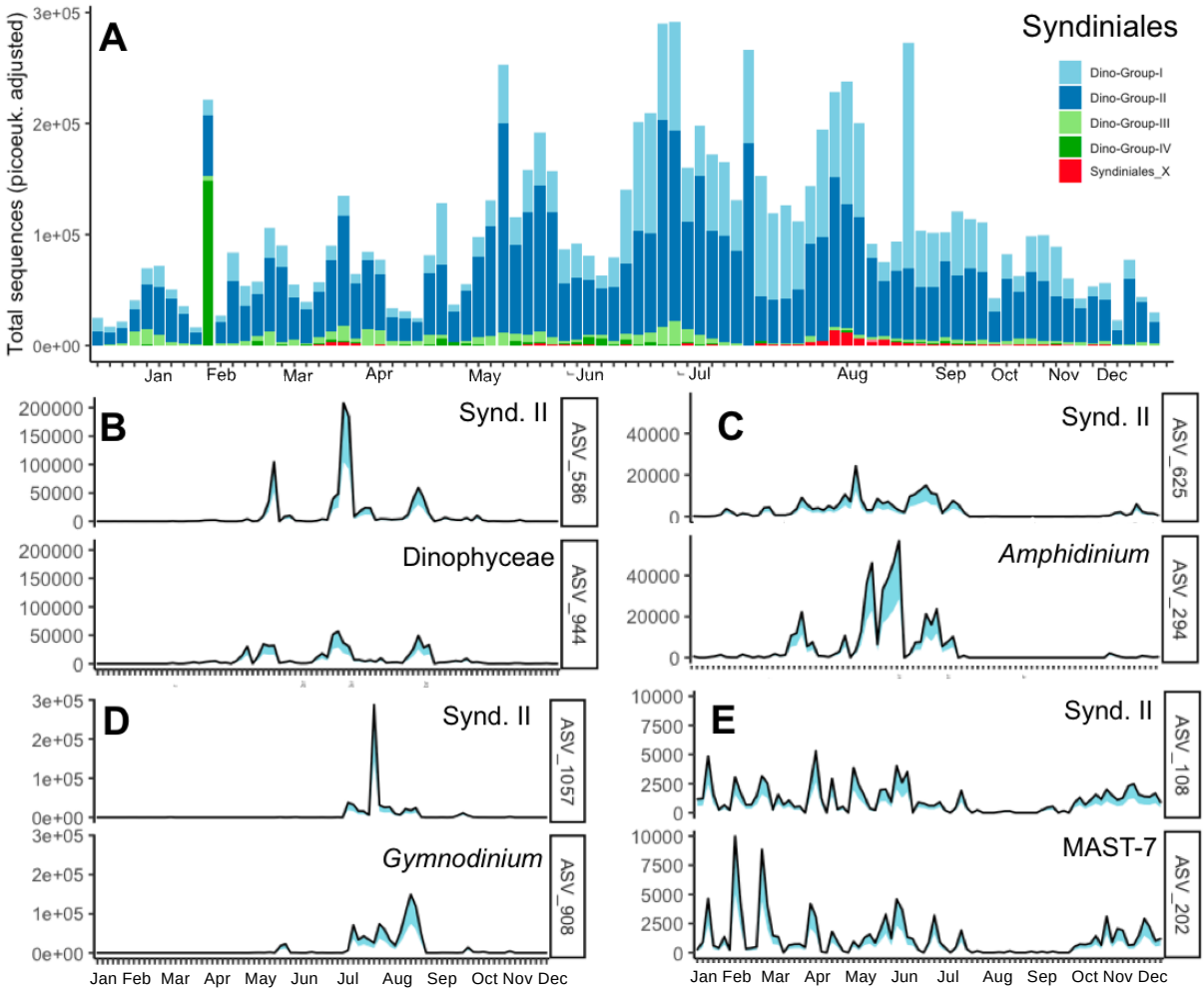


Figure 3.6 (A) Adjusted Syndiniales counts throughout 2016 (B-E) Adjusted counts of four putative parasite-host pairs that had significant correlations using SparCC (B) r = 0.74 (C) r = 0.625 (D) r = 0.71 (E) r = 0.74.

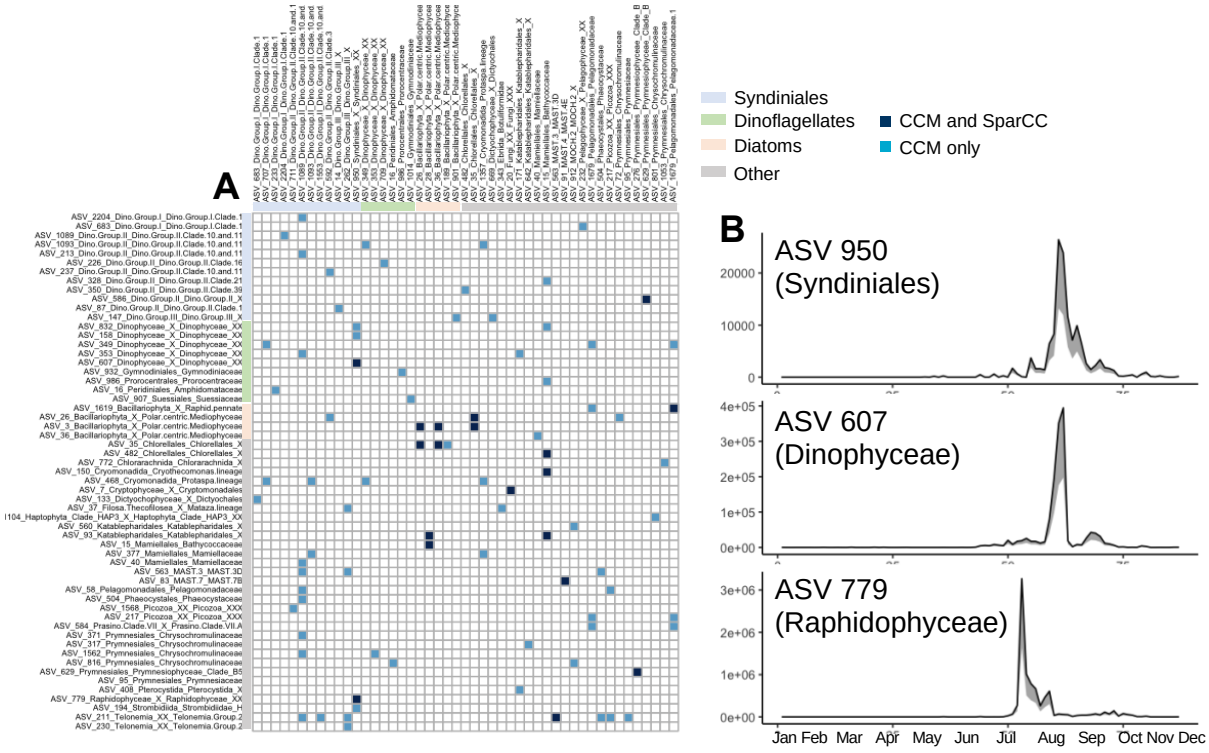


Figure 3.7 (A) CCM matrix of ASV pairs with significant causal interactions. Causality is unidirectional and in this matrix should be interpreted as the column being causal to the row. Taxa are grouped and colors under ASV names indicate taxonomic groupings. Colors of boxes indicate whether each interaction was significant using CCM only or both CCM and SparCC. (B) ASVs which had significant SparCC correlations along with causal signature. The Syndiniales ASV has a causal relationship with both of the others.

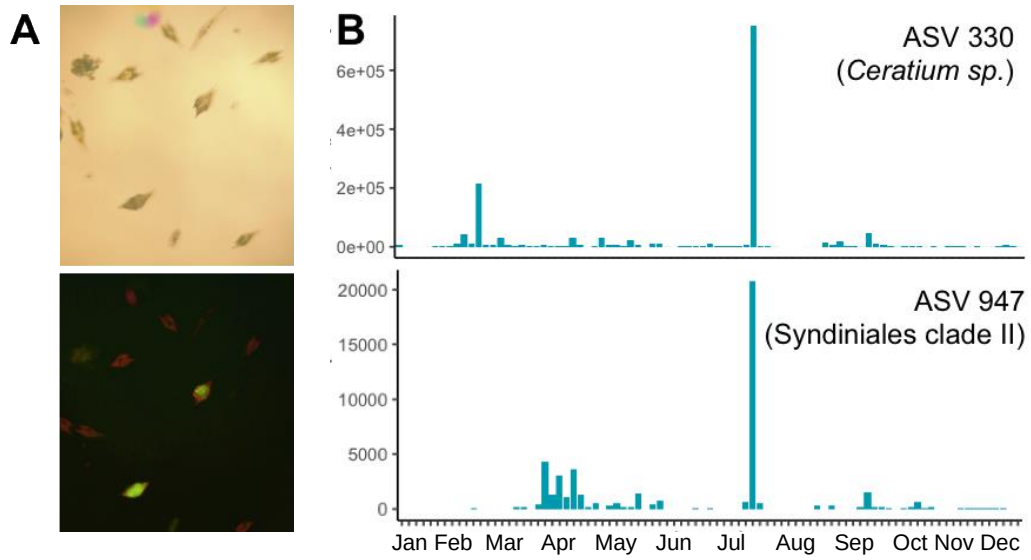


Figure 3.8 (A) March 2018 bloom with many *Ceratium* sp. present in bright field (top) and under blue light (bottom). Red fluorescence is the chloroplasts and green fluorescence indicates *Amoebophrya* infection. (B) Putative ASVs representing the dinoflagellate-host pair on this date and their abundance in the 2016 time series.

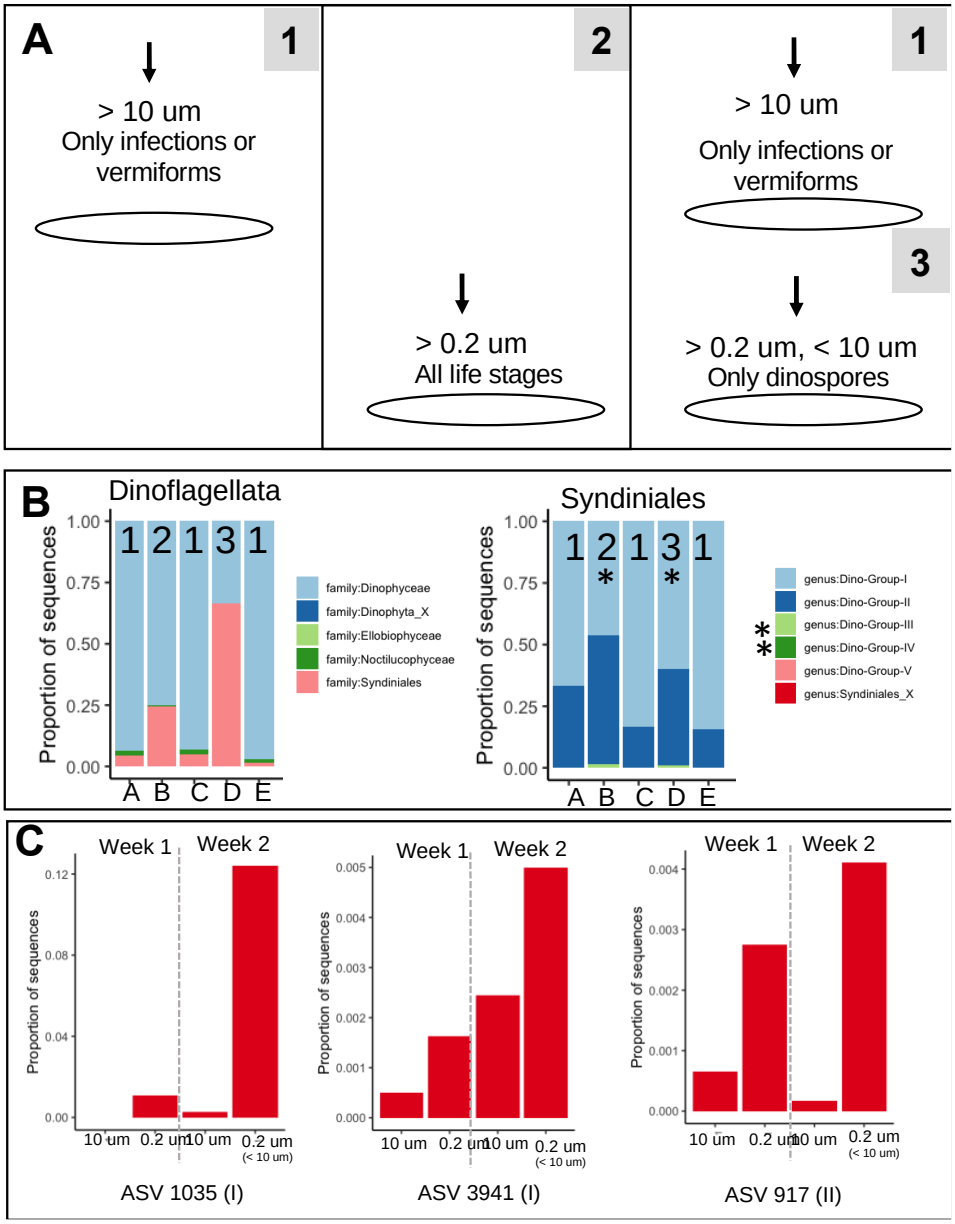


Figure 3.9 (A) Schematic representation of filtration procedure for different samples shown (B) Relative taxonomic abundances of of all dinoflagellate (top) and all Syndiniales (bottom) ASVs amongst different filter size fractions. Numbers designate the filtration method from (A). Letters represent date sampled; A – 10 July 2017; B & C – 17 July 2017; D & E – 27 July 2017. Asterisks indicate that the sample contained sequences from groups III and IV. (C) Comparison of three Syndiniales ASVs filtered at two size fractions each, one week apart.

3.7 ACKNOWLEDGEMENTS

This research was funded by a UCSD Academic Senate Grant and computational resources were provided by an XSEDE allocation (project OCE180001). We thank Tsvetan Bachvaroff for sharing his *Karlodinium* and *Amoebophrya* cultures and providing important insights on their dynamics. We thank Ryan Hechinger for his many thoughts on parasitism and ecology, George Sugihara for his direction in time series analysis, and Karsten Zengler for his suggestions about wrangling sequence data. We also thank Chase James, Gerald Pao, and Alfredo Giron-Nava for advice regarding data analysis. Finally, we thank Maggie Wang for her assistance with some of the sample collection and DNA extractions.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Nagarkar, Maitreyi; Palenik, Brian. The dissertation author was the primary author on this paper.

3.8 REFERENCES

Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W., & Huse, S. M. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA Genes. *PLoS ONE*, *4*(7), 1–9. https://doi.org/10.1371/journal.pone.0006372

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., et al. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *MSystems*, *2*(2), 1–7.

Anderson, R. M., & May, R. M. (1978). Regulation and Stability of Host-Parasite Population Interactions. *Journal of Animal Ecology*, *47*(1), 219–247. https://doi.org/10.2307/3933

Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. a, & Thingstad, F. (1983). The Ecological Role of Water-Column Microbes in the Sea. *Marine Ecology Progress Series*. https://doi.org/10.3354/meps010257

Bai, X., Adolf, J. E., Bachvaroff, T., Place, A. R., & Coats, D. W. (2007). The interplay between host toxins and parasitism by Amoebophrya. *Harmful Algae*, *6*(5), 670–678. https://doi.org/10.1016/j.hal.2007.01.003

Bjorbækmo, M. F. M., Evenstad, A., Røsæg, L. L., & Anders, K. (2019). The planktonic protist interactome: where do we stand after a century of research? *BioRxiv*. https://doi.org/http://dx.doi.org/10.1101/587352

Chambouvet, A., Morin, P., Marie, D., & Guillou, L. (2008). Control of Toxic Marine Dinoflagellate Blooms by Serial Parasitic Killers. *Science*, *322*(5905), 1254–1257. https://doi.org/10.1126/science.1164387

Chambouvet, Aurélie, Laabir, M., Sengco, M., Vaquer, A., & Guillou, L. (2011). Genetic diversity of Amoebophryidae (Syndiniales) during *Alexandrium catenella/tamarense* (Dinophyceae) blooms in the Thau lagoon (Mediterranean Sea, France). *Research in Microbiology*, *162*(9), 960–968. https://doi.org/10.1016/j.resmic.2011.03.002

Cleary, A. C., & Durbin, E. G. (2016). Unexpected prevalence of parasite 18S rDNA sequences in winter among Antarctic marine protists. *Journal of Plankton Research*, *0*(0), 1–17. https://doi.org/10.1093/plankt/fbw005

Coats, D. W., Adam, E. J., Gallegos, C. L., & Hedrick, S. (1996). Parasitism of photosynthetic dinoflagellates in a shallow subestuary of Chesapeake Bay, USA. *Aquatic Microbial Ecology*, *11*(1), 1–9. https://doi.org/10.3354/ame011001

Coats, D. Wayne, & Bockstahler, K. R. (1994). Occurrence of the Parasitic Dinoflagellate

Amoebophrya ceratii in Chesapeake Bay Populations of *Gymnodinium sanguineum*. *Journal of Eukaryotic Microbiology*, *41*(6), 586–593.

Coats, D. Wayne, Kim, S., Bachvaroff, T. R., Handy, S. M., & Delwiche, C. F. (2010). Tintinnophagus acutus n. g., n. sp. (phylum dinoflagellata), an ectoparasite of the ciliate Tintinnopsis cylindrica Daday 1887, and its relationship to Duboscquodinium collini Grassé 1952. *Journal of Eukaryotic Microbiology*, *57*(6), 468–482. https://doi.org/10.1111/j.1550-7408.2010.00504.x

Coats, D W. (1999). Parasitic life styles of marine dinoflagellates. *Journal of Eukaryotic Microbiology*, *46*(4), 402–409. https://doi.org/10.1111/j.1550-7408.1999.tb04620.x

Coats, D Wayne, & Park, M. G. (2002). Parasitism of photosynthetic dinoflagellates by three strains of. *Aquatic Microbial Ecology*, *528*, 520–528. https://doi.org/10.1046/j.1529-8817.2002.01200.x

Friedman, J., & Alm, E. J. (2012). Inferring Correlation Networks from Genomic Survey Data. *PLoS Computational Biology*, *8*(9), 1–11. https://doi.org/10.1371/journal.pcbi.1002687

Gao, X., Lin, H., Revanna, K., & Dong, Q. (2017). A Bayesian taxonomic classification method for 16S rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics*, *18*(1), 1–10. https://doi.org/10.1186/s12859-017-1670-4

Giron-Nava, A., James, C. C., Johnson, A. F., Dannecker, D., Kolody, B., Lee, A., *et al*. (2017). Quantitative argument for long-term ecological monitoring. *Marine Ecology Progress Series*, *572*, 269–274. https://doi.org/10.3354/meps12149

Guillou, L., Viprey, M., Chambouvet, A., Welsh, R. M., Kirkham, A. R., Massana, R., et al. (2008a). Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental Microbiology*, *10*(12), 3349–3365. https://doi.org/10.1111/j.1462-2920.2008.01731.x

Guillou, L., Viprey, M., Chambouvet, A., Welsh, R. M., Kirkham, A. R., Massana, R., et al. (2008b). Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environmental Microbiology*, *10*(12), 3349–3365. https://doi.org/10.1111/j.1462-2920.2008.01731.x

Guillou, Laure, Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., et al. (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*(D1), 597–604. https://doi.org/10.1093/nar/gks1160

Johansson, M., & Coats, D. W. (2002). Ciliate grazing on the parasite *Amoebophrya sp.* decreases infection of the red-tide dinoflagellate *Akashiwo sanguinea*. *Aquatic Microbial Ecology*, *28*(1), 69–78. https://doi.org/Doi 10.3354/Ame028069

Jung, J. H., Choi, J. M., Coats, D. W., & Kim, Y. O. (2016). *Euduboscquella costata* n. sp. (Dinoflagellata, Syndinea), an Intracellular Parasite of the Ciliate *Schmidingerella arcuata*: Morphology, Molecular Phylogeny, Life Cycle, Prevalence, and Infection Intensity. *Journal of Eukaryotic Microbiology*, *63*(1), 3–15. https://doi.org/10.1111/jeu.12231

Kim, S., Park, M. G., Kim, K. Y., Kim, C. H., Yih, W., Park, J. S., & Coats, D. W. (2008). Genetic diversity of parasitic dinoflagellates in the genus Amoebophrya and its relationship to parasite biology and biogeography. *Journal of Eukaryotic Microbiology*, *55*(1), 1–8. https://doi.org/10.1111/j.1550-7408.2007.00295.x

Kuris, A. M. (1974). Trophic Interactions: Similarity of Parasitic Castrators to Parasitoids. *The Quarterly Review of Biology*, *49*, 129–148.

Lima-mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., et al. (2015). Determinants of community structure in the global plankton interactome. *Science*, *348*(6237), 1–10.

Majaneva, M., Rintala, J. M., Piisilä, M., Fewer, D. P., & Blomster, J. (2012). Comparison of wintertime eukaryotic community from sea ice and open water in the Baltic Sea, based on sequencing of the 18S rRNA gene. *Polar Biology*, *35*(6), 875–889. https://doi.org/10.1007/s00300-011-1132-9

Mazzillo, F. F. M., Ryan, J. P., & Silver, M. W. (2011). Parasitism as a biological control agent of dinoflagellate blooms in the California Current System. *Harmful Algae*, *10*(6), 763–773. https://doi.org/10.1016/j.hal.2011.06.009

Mcmurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061217

Montagnes, D. J. S., Chambouvet, A., Guillou, L., & Fenton, A. (2008). Responsibility of microzooplankton and parasite pressure for the demise of toxic dinoflagellate blooms. *Aquatic Microbial Ecology*, *53*(2), 211–225. https://doi.org/10.3354/ame01245

Moon-van der Staay, S. Y., De Wachter, R., & Vaulot, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature*, *409*(6820), 607–610. https://doi.org/10.1038/35054541

Nagarkar, M., Countway, P. D., Du, Y., Daniels, E., Poulton, N. J., & Palenik, B. (2018). Temporal dynamics of eukaryotic microbial diversity at a coastal Pacific site. *The ISME Journal*, 2278–2291. https://doi.org/10.1038/s41396-018-0172-3

Park, M. G., Kim, S., Shin, E.-Y., Yih, W., & Coats, D. W. (2013). Parasitism of harmful dinoflagellates in Korean coastal waters. *Harmful Algae*, *30*, S62–S74. https://doi.org/http://dx.doi.org/10.1016/j.hal.2013.10.007

Satinsky, B. M., Gifford, S. M., Crump, B. C., & Moran, M. A. (2013). Use of Internal Standards for Quantitative Metatranscriptome and Metagenome Analysis. Microbial Metagenomics, Metatranscriptomics, and Metaproteomics (1st ed., Vol. 531). Elsevier Inc. https://doi.org/10.1016/B978-0-12-407863-5.00012-5

Shields, J. D. (1994). The parasitic dinofalgellates of marine crustaceans. *Annual Review of Fish Diseases*, *4*, 241–271.

Siano, R., Alves-De-Souza, C., Foulon, E., M. Bendif, E., Simon, N., Guillou, L., & Not, F. (2011). Distribution and host diversity of Amoebophryidae parasites across oligotrophic waters of the Mediterranean Sea. *Biogeosciences*, *8*(2), 267–278. https://doi.org/10.5194/bg-8-267-2011

Skovgaard, A., Massana, R., Balague, V., & Saiz, E. (2005). Phylogenetic position of the copepod-infesting parasite *Syndinium turbo* (Dinoflagellata, Syndinea). *Protist*, *156*(4), 413–423. https://doi.org/10.1016/j.protis.2005.08.002

Skovgaard, A., Daugbjerg, N., Chambouvet, A., Alves-de-Souza, C., Cueff, V., Marie, D., et al. (2008). Parasitism of photosynthetic dinoflagellates by three strains of *Amoeobophrya* (Dinophyta): Parasite survival, infectivity, generation time, and host specificity. *Protist*, *66*(1), 1254–1257. https://doi.org/10.1093/plankt/fbw005

Stentiford, G. D., & Shields, J. D. (2005). A review of the parasitic dinoflagellates *Hematodinium* species and *Hematodinium*-like infections in marine crustaceans. *Diseases of Aquatic Organisms*, *66*(1), 47–70. https://doi.org/10.3354/dao066047

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H. W., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, *19*, 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Munch, S., et al. (2012). Detecting Causality in Complex Ecosystems. *Science*, *338*(6106), 496–500.

Taylor, F. J. R. (1968). Parasitism of the toxin-producing dinoflagellate Gonyaulax catenella by the endoparasitic dinoflagellate *Amoebophrya ceratii*. J*ournal of the Fisheries Board of Canada*, 25(10), 2241-2245.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237), 1261605. https://doi.org/10.1126/science.1261605

Velo-Suárez, L., Brosnahan, M. L., Anderson, D. M., & McGillicuddy, D. J. (2013). A quantitative assessment of the role of the parasite *Amoebophrya* in the termination of *Alexandrium fundyense* blooms within a small coastal embayment. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0081150

Wang, S., Lin, Y., Gifford, S., Eveleth, R., & Cassar, N. (2018). Linking patterns of net community production and marine microbial community structure in the western North Atlantic. *The ISME Journal*, 2582–2595. https://doi.org/10.1038/s41396-018-0163-4

Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., et al. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME Journal*, *10*(7), 1669–1681. https://doi.org/10.1038/ismej.2015.235

Ye, H., Beamish, R. J., Glaser, S. M., Grant, S. C. H., Hsieh, C., Richards, L. J., et al. (2015). Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proceedings of the National Academy of Sciences*, *112*(13), E1569–E1576. https://doi.org/10.1073/pnas.1417063112

APPENDIX

Supplementary text S1. Illumina 2-step protocol from RTL Genomics

Samples were amplified for sequencing in a two-step process. The forward primer was constructed with (5'-3') the Illumina i5 sequencing primer (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) and the Euk_1391F primer (Amaral-Zettler et al., 2009). The reverse primer was constructed with (5'-3') the Illumina i7 sequencing primer (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) and the EukBr primer (Stoeck et al., 2010). Amplifications were performed in 25 µl reactions with Qiagen HotStar Taq master mix (Qiagen Inc, Valencia, California), 1ul of each 5µM primer, and 1µl of template. Reactions were performed on ABI Veriti thermocyclers (Applied Biosystems, Carlsbad, California) under the following thermal profile: 95°C for 5 min, then 35 cycles of 94°C for 30 s, 54°C for 40 s, 72°C for 1 min, followed by one cycle of 72°C for 10 min and 4°C hold. Products from the first stage amplification were added to a second PCR based on qualitatively determine concentrations. Primers for the second PCR were designed based on the Illumina Nextera PCR primers as follows: Forward - AATGATACGGCGACCACCGAGATCTACAC[i5index]TCGTCGGCAGCGTC and Reverse - CAAGCAGAAGACGGCATACGAGAT[i7index]GTCTCGTGGGCTCGG. The second stage amplification was run the same as the first stage except for 10 cycles.

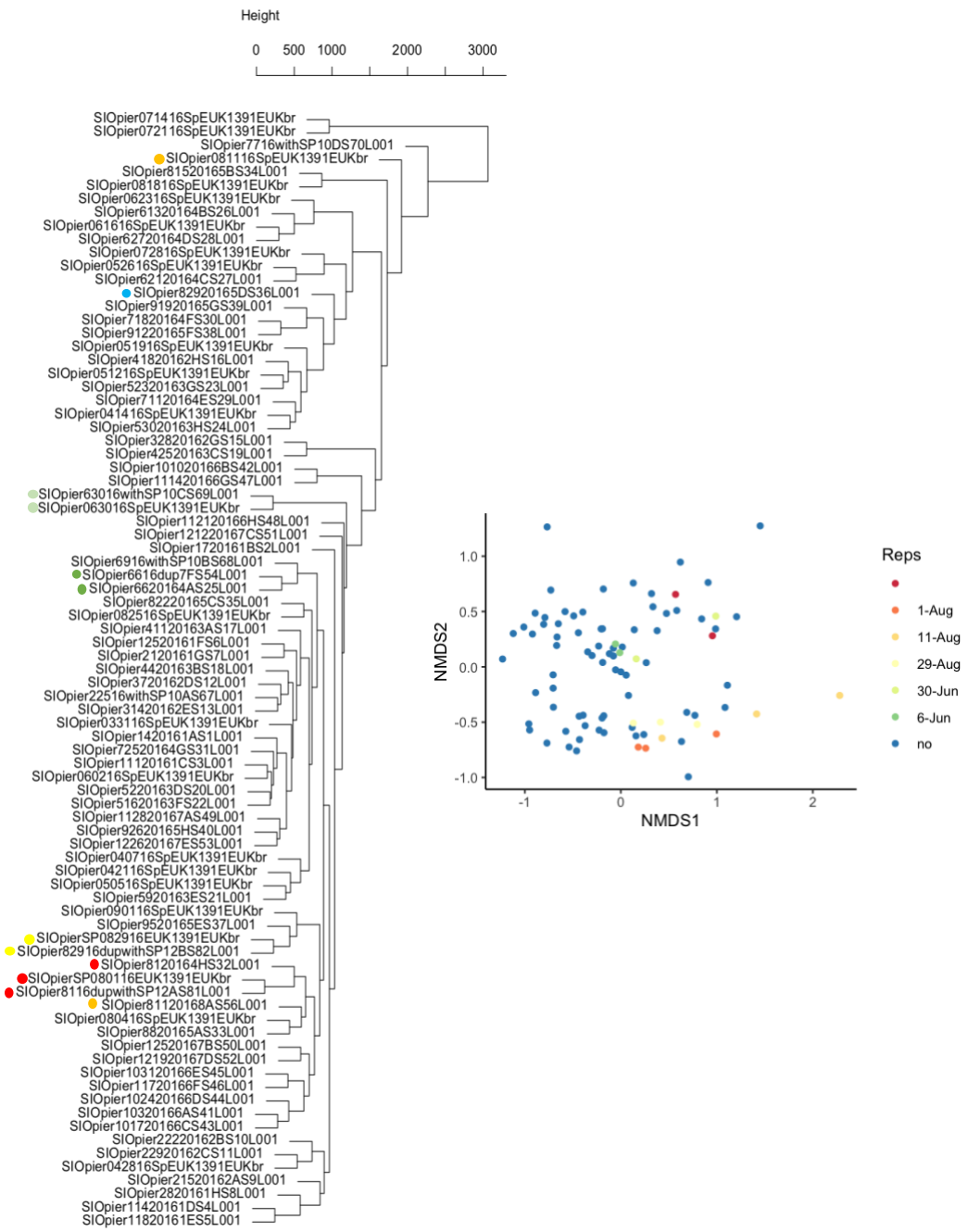
Amplification products were visualized with eGels (Life Technologies, Grand Island, New York). Products were then pooled equimolar and each pool was size selected in two rounds using SPRIselect Reagent (Beckman Coulter, Indianapolis, Indiana) in a 0.75 ratio for both rounds. Size selected pools were then quantified using the Qubit 4 Fluorometer (Life Technologies) and loaded on an Illumina MiSeq (Illumina, Inc. San Diego, California) 2x300 flow cell at 10pM.

Supplementary text S2. Use of *S. pombe* internal standard

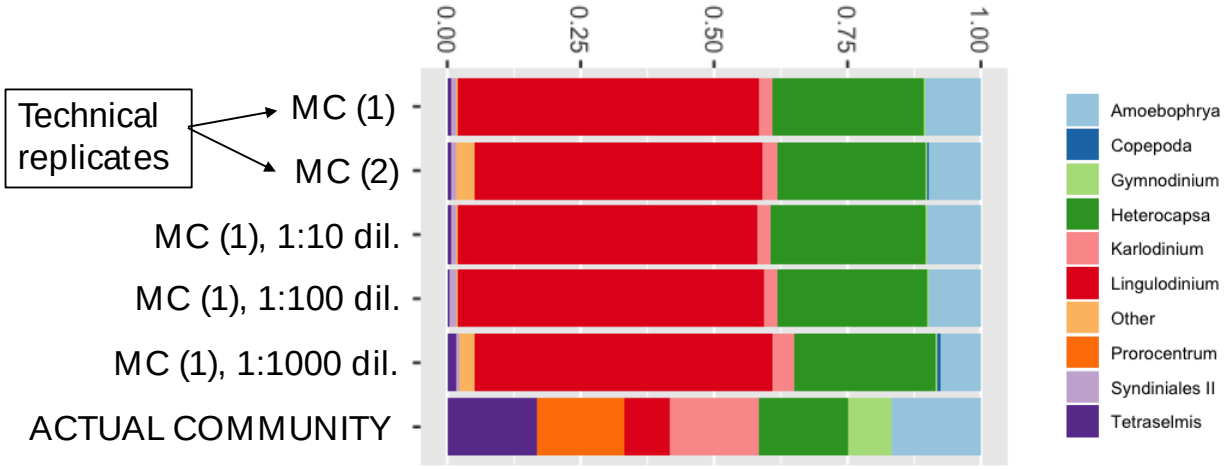
Initially we intended to use an internal DNA standard to normalize counts of DNA as described in several previous studies (Satinsky et al., 2013; Wang et al., 2018). *S. pombe* Lindner genomic DNA obtained from ATCC (Manassas, VA, USA) was re-hydrated in 1 ml sterile water and concentrations of subsequent dilutions were confirmed on a Qubit fluorometer. We added 0.1 ng of *S. pombe* DNA to each sample either before or after DNA extraction; when added before DNA extraction it was added directly to the tubes that cut filters were subsequently put in.

However, we ultimately chose to remove all *S. pombe* sequences and normalize using flow cytometric counts of photosynthetic picoeukaryotes. This method was chosen over the use of the *S. pombe* internal standard for two reasons: firstly, the number of *S. pombe* sequences was too low for the majority of samples where it was added because of low sequencing depth, including one sample where *S. pombe* had been added but yielded no sequences at a depth of 10K. Because of this we did not feel confident in our adjusted values for these samples. Secondly, the relative numbers of *Mamiellales* sequences were not consistent with the patterns observed in flow cytometry counts. We have higher confidence in the flow cytometry counts because these correlated well with both the relative abundances and a similar adjustment made using *Prorocentrum spp.* counts (from microscopy).

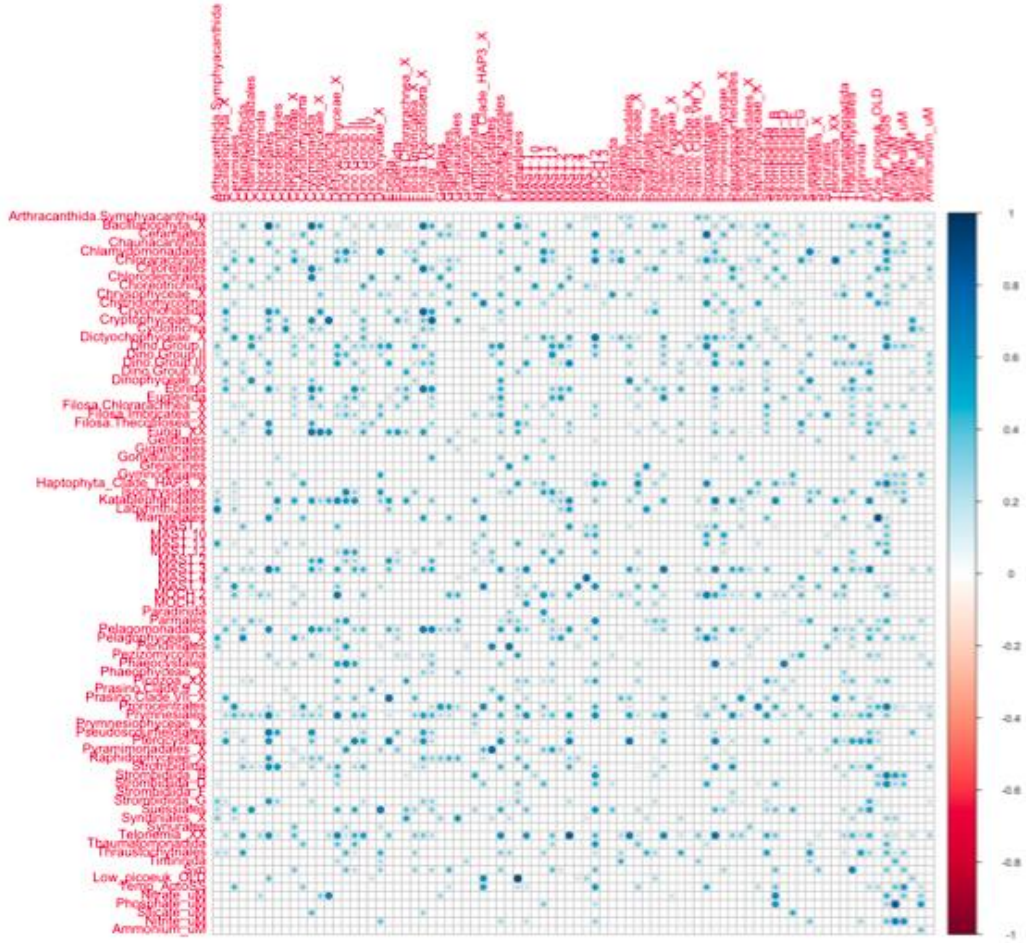
While of the picoeukaryote data to adjust sequences did change the temporal patterns of individual ASVs, the mean Pearson's correlation between the relative proportion of all ASVs and the adjusted counts of all ASVs (after any with a correlation of 1 are removed) was 0.91. Thus it appears that there were not strong compositional effects on our particular data. Nevertheless, it is possible that ASVs of interest do experience strong effects so quantification is still a worthwhile effort.



Supplementary figure S3. Clustering of samples with an emphasis on similarity of replicates. Left: Hierarchical clustering of all SIO pier samples, with replicates indicated as the same color. Names that include "L001" were sequenced at the IGM facility while all others were sequenced at RTL Genomics; three sets of replicates sequenced at different facilities still clustered together. Right: NMDS plot with replicates specified. *S. pombe* sequences have been removed from all samples.



Supplementary figure S4. 18S sequence results of eukaryotic mock communities. Actual proportions of cells added shown in the bottom row; all other rows represent proportions of sequences found in various replicates. Several of the communities contain contaminants ("Other" and "Copepod" sequences).



Supplementary figure S5. Results of CCM using counts aggregated by broader taxa, along with several environmental variables: temperature, nitrate, phosphate, silicate, nitrite, and ammonium. Direction of causality is "column causes row" and size of the circle represents the rho value on the sliding scale.

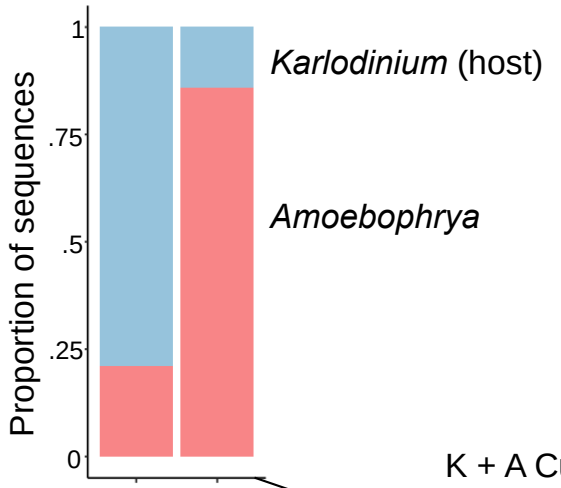
Supplementary table S6. Pairs of ASVs that had a significant causal interaction (column 1 causes column 2) identified using CCM (with both sets of time series) and also had a significant ($p<0.05$) correlation using SparCC. Pairs that include a Syndiniales ASV are bolded.

ASV #1		ASV #2		Correlation	Found in lit.?
ASV_842	Dino-Group-I	ASV_779	Raphidophyceae	0.77976319	
ASV_842	Dino-Group-I	ASV_607	Dinophyceae	0.74571651	
ASV_586	Dino-Group-II	ASV_944	Dinophyceae	0.7408323	x
ASV_950	Syndiniales_X	ASV_900	Pyramimonadales	0.73613077	
ASV_108	Dino-Group-II	ASV_202	MAST	0.73508926	
ASV_842	Dino-Group-I	ASV_951	Telonemia_X	0.73478565	
ASV_237	Dino-Group-II	ASV_28	Bacillariophyta	0.72749677	
ASV_842	Dino-Group-I	ASV_961	Dinophyceae	0.72355863	
ASV_842	Dino-Group-I	ASV_950	Syndiniales	0.71557982	
ASV_950	Syndiniales_X	ASV_842	Syndiniales	0.71557982	
ASV_842	Dino-Group-I	ASV_1037	Dinophyceae	0.71478096	
ASV_2204	Dino-Group-I	ASV_1711	Filosa-Thecofilosea	0.71291377	x
ASV_108	Dino-Group-II	ASV_25	Mamiellophyceae	0.71018024	
ASV_958	Dino-Group-II	ASV_912	MOCH	0.70934015	
ASV_842	Dino-Group-I	ASV_900	Pyramimonadales	0.70808479	
ASV_1057	Dino-Group-II	ASV_908	Dinophyceae	0.70717388	x
ASV_328	Dino-Group-II	ASV_28	Bacillariophyta	0.70677421	
ASV_950	Syndiniales_X	ASV_779	Raphidophyceae	0.70665937	
ASV_950	Syndiniales_X	ASV_1037	Dinophyceae	0.70086009	
ASV_237	Dino-Group-II	ASV_3	Bacillariophyta	0.69828981	
ASV_842	Dino-Group-I	ASV_933	Spirotrichea	0.69705955	
ASV_14	Dino-Group-III	ASV_25	Mamiellophyceae	0.69410637	
ASV_950	Syndiniales_X	ASV_951	Telonemia_X	0.69304318	
ASV_842	Dino-Group-I	ASV_1014	Dinophyceae	0.69289397	
ASV_328	Dino-Group-II	ASV_1711	Filosa-Thecofilosea	0.69274637	
ASV_950	Syndiniales_X	ASV_1014	Dinophyceae	0.69203009	
ASV_14	Dino-Group-III	ASV_171	Katablepharidaceae	0.69146688	
ASV_237	Dino-Group-II	ASV_1711	Filosa-Thecofilosea	0.68930233	
ASV_625	Dino-Group-II	ASV_36	Bacillariophyta	0.68750246	
ASV_1057	Dino-Group-II	ASV_779	Raphidophyceae	0.6860556	
ASV_950	Syndiniales_X	ASV_2086	Spirotrichea	0.6850761	

ASV_2160	Dino-Group-I	ASV_28	Bacillariophyta	0.6843221	
ASV_586	Dino-Group-II	ASV_232	Pelagophyceae	0.68273832	
ASV_950	Syndiniales_X	ASV_967	Spirotrichea	0.6808935	
ASV_14	Dino-Group-III	ASV_40	Mamiellophyceae	0.68064002	
ASV_950	Syndiniales_X	ASV_933	Spirotrichea	0.67938911	
ASV_964	Dino-Group-I	ASV_779	Raphidophyceae	0.67937842	
ASV_950	Syndiniales_X	ASV_961	Dinophyceae	0.67840046	
ASV_328	Dino-Group-II	ASV_35	Trebouxiophyceae	0.67472032	
ASV_950	Syndiniales_X	ASV_607	Dinophyceae	0.67238055	
ASV_182	Dino-Group-II	ASV_35	Trebouxiophyceae	0.67213046	
ASV_842	Dino-Group-I	ASV_1077	Oomycota	0.67198048	
ASV_586	Dino-Group-II	ASV_629	Prymnesiophyceae	0.66859694	
ASV_929	Dino-Group-II	ASV_942	Syndiniales	0.66858806	
ASV_942	Dino-Group-II	ASV_929	Syndiniales	0.66858806	
ASV_1057	Dino-Group-II	ASV_933	Spirotrichea	0.66727516	x
ASV_237	Dino-Group-II	ASV_2204	Syndiniales	0.66663263	
ASV_2204	Dino-Group-I	ASV_237	Syndiniales	0.66663263	
ASV_2160	Dino-Group-I	ASV_1570	Dinophyceae	0.66463359	
ASV_291	Dino-Group-II	ASV_482	Trebouxiophyceae	0.66393781	
ASV_950	Syndiniales_X	ASV_1054	Telonemia_X	0.65894545	
ASV_973	Dino-Group-II	ASV_2044	Polycystinea	0.65761837	
ASV_929	Dino-Group-II	ASV_924	Pyramimonadales	0.65683844	
ASV_842	Dino-Group-I	ASV_2086	Spirotrichea	0.65592388	
ASV_586	Dino-Group-II	ASV_866	Chlorophyceae	0.65466988	
ASV_182	Dino-Group-II	ASV_28	Bacillariophyta	0.65446986	
ASV_237	Dino-Group-II	ASV_35	Trebouxiophyceae	0.65356887	
ASV_625	Dino-Group-II	ASV_58	Pelagophyceae	0.65339831	
ASV_625	Dino-Group-II	ASV_294	Dinophyceae	0.65239243	x
ASV_929	Dino-Group-II	ASV_560	Katablepharidaceae	0.65210347	
ASV_1013	Dino-Group-II	ASV_1014	Dinophyceae	0.65189268	x
ASV_328	Dino-Group-II	ASV_3	Bacillariophyta	0.65126837	
ASV_842	Dino-Group-I	ASV_967	Spirotrichea	0.6506703	
ASV_195	Dino-Group-I	ASV_32	Dinophyceae	0.6506509	
ASV_842	Dino-Group-I	ASV_1054	Telonemia_X	0.64971416	
ASV_842	Dino-Group-I	ASV_898	Syndiniales	0.6492194	
ASV_898	Dino-Group-II	ASV_842	Syndiniales	0.6492194	
ASV_964	Dino-Group-I	ASV_951	Telonemia_X	0.64896644	

ASV_2160	Dino-Group-I	ASV_1711	Filosa-Thecofilosea	0.64864718	x
ASV_964	Dino-Group-I	ASV_1037	Dinophyceae	0.64847268	
ASV_237	Dino-Group-II	ASV_26	Bacillariophyta	0.64141535	
ASV_929	Dino-Group-II	ASV_772	Filosa-Chlorarachnea	0.64109134	
ASV_929	Dino-Group-II	ASV_945	Chrysophyceae	0.64097194	
ASV_237	Dino-Group-II	ASV_2	Bacillariophyta	0.64064412	
ASV_2204	Dino-Group-I	ASV_708	Filosa-Thecofilosea	0.64054695	x
ASV_2204	Dino-Group-I	ASV_3	Bacillariophyta	0.64048188	
ASV_1035	Dino-Group-I	ASV_1093	Syndiniales	0.64033777	
ASV_1093	Dino-Group-II	ASV_1035	Syndiniales	0.64033777	
ASV_880	Dino-Group-I	ASV_779	Raphidophyceae	0.64000287	
ASV_1013	Dino-Group-II	ASV_779	Raphidophyceae	0.63968052	
ASV_182	Dino-Group-II	ASV_3	Bacillariophyta	0.63945711	
ASV_163	Dino-Group-I	ASV_36	Bacillariophyta	0.63945611	
ASV_108	Dino-Group-II	ASV_40	Mamiellophyceae	0.63934067	
ASV_108	Dino-Group-II	ASV_22	Pelagophyceae	0.63688496	
ASV_842	Dino-Group-I	ASV_848	Spirotrichea	0.63364025	
ASV_842	Dino-Group-I	ASV_1041	MAST	0.63202525	
ASV_182	Dino-Group-II	ASV_1711	Filosa-Thecofilosea	0.63138782	
ASV_1035	Dino-Group-I	ASV_983	Trebouxiophyceae	0.63114136	
ASV_942	Dino-Group-II	ASV_924	Pyramimonadales	0.62991975	
ASV_592	Dino-Group-II	ASV_28	Bacillariophyta	0.62735275	
ASV_950	Syndiniales_X	ASV_1222		0.62690132	
ASV_2204	Dino-Group-I	ASV_35	Trebouxiophyceae	0.62596972	
ASV_49	Dino-Group-III	ASV_84	Dinophyceae	0.62519009	
ASV_1013	Dino-Group-II	ASV_877	Arthropoda	0.62440037	
ASV_964	Dino-Group-I	ASV_1041	MAST	0.62343466	
ASV_14	Dino-Group-III	ASV_133	Dictyochophyceae	0.62319308	
ASV_950	Syndiniales_X	ASV_1041	MAST	0.62295337	
ASV_195	Dino-Group-I	ASV_150	Filosa-Thecofilosea	0.62212648	x
ASV_950	Syndiniales_X	ASV_1317	Syndiniales	0.62211797	
ASV_1317	Dino-Group-I	ASV_950	Syndiniales	0.62211797	
ASV_929	Dino-Group-II	ASV_1035	Syndiniales	0.62009366	
ASV_1035	Dino-Group-I	ASV_929	Syndiniales	0.62009366	
ASV_57	Dino-Group-II	ASV_103	Syndiniales	0.61831196	
ASV_103	Dino-Group-II	ASV_57	Syndiniales	0.61831196	
ASV_842	Dino-Group-I	ASV_996	Syndiniales	0.61793144	
ASV_996	Dino-Group-II	ASV_842	Syndiniales	0.61793144	

ASV_163	Dino-Group-I	ASV_2	Bacillariophyta	0.6172603	
ASV_1271	Dino-Group-II	ASV_779	Raphidophyceae	0.61641086	
ASV_586	Dino-Group-II	ASV_560	Katablepharidaceae	0.6155682	
ASV_1317	Dino-Group-I	ASV_607	Dinophyceae	0.61530388	
ASV_1271	Dino-Group-II	ASV_933	Spirotrichea	0.61487936	x
ASV_1057	Dino-Group-II	ASV_1020	Prymnesiophyceae	0.61482116	
ASV_57	Dino-Group-II	ASV_31	Spirotrichea	0.6146264	x
ASV_1057	Dino-Group-II	ASV_607	Dinophyceae	0.61460837	x
ASV_57	Dino-Group-II	ASV_195	Syndiniales	0.61455247	
ASV_195	Dino-Group-I	ASV_57	Syndiniales	0.61455247	
ASV_108	Dino-Group-II	ASV_218	MAST	0.61319492	
ASV_1317	Dino-Group-I	ASV_951	Telonemia_X	0.61302045	
ASV_586	Dino-Group-II	ASV_373	Dinophyceae	0.61264195	x
ASV_182	Dino-Group-II	ASV_98		0.61243179	
ASV_942	Dino-Group-II	ASV_993	MAST	0.61229603	
ASV_328	Dino-Group-II	ASV_281	Bacillariophyta	0.61118052	
ASV_57	Dino-Group-II	ASV_2	Bacillariophyta	0.61051205	
ASV_929	Dino-Group-II	ASV_983	Trebouxiophyceae	0.61022591	
ASV_2160	Dino-Group-I	ASV_3	Bacillariophyta	0.60995779	
ASV_57	Dino-Group-II	ASV_91	MAST	0.60920715	
ASV_109	Dino-Group-II	ASV_11	Dinophyceae	0.60919959	x
ASV_237	Dino-Group-II	ASV_58	Pelagophyceae	0.60896322	
ASV_880	Dino-Group-I	ASV_837	Filosa-Chlorarachnea	0.6086405	x
ASV_842	Dino-Group-I	ASV_67	Bacillariophyta	0.60837845	
ASV_237	Dino-Group-II	ASV_150	Filosa-Thecofilosea	0.60814088	
ASV_249	Dino-Group-II	ASV_204	Ascomycota	0.60780639	
ASV_109	Dino-Group-II	ASV_101		0.60759206	
ASV_109	Dino-Group-II	ASV_25	Mamiellophyceae	0.6070434	
ASV_249	Dino-Group-II	ASV_137	Bacillariophyta	0.60696427	
ASV_328	Dino-Group-II	ASV_482	Trebouxiophyceae	0.60666077	
ASV_103	Dino-Group-II	ASV_291	Syndiniales	0.60654446	
ASV_291	Dino-Group-II	ASV_103	Syndiniales	0.60654446	
ASV_707	Dino-Group-I	ASV_95	Prymnesiophyceae	0.60596081	
ASV_2160	Dino-Group-I	ASV_35	Trebouxiophyceae	0.60579533	
ASV_108	Dino-Group-II	ASV_15	Mamiellophyceae	0.6054881	
ASV_950	Syndiniales_X	ASV_1077	Oomycota	0.60271761	
ASV_1317	Dino-Group-I	ASV_779	Raphidophyceae	0.60181015	
ASV_973	Dino-Group-II	ASV_1443	Chrysophyceae	0.60093294	
ASV_1013	Dino-Group-II	ASV_961	Dinophyceae	0.60016888	x



K + A Culture
>24,950,000 cells

Healthy *Karlo.*	2,000,000
Early st infection	160,000
Mid st infection	430,000
Late st infection	160,000
Dinospores	22,200,000

K + A Mock Comm.
1,400,000 cells

Healthy *Karlo.*	700,000
Dinospores	700,000

Supplementary figure S7. 18S sequence results of two mock communities made with only *Karlodinium* (host) and *Amoebophrya* (Syndiniales parasite) cells. When added in equal proportions, *Karlodinium* clearly has a higher 18S copy number than *Amoebophrya* and is over-represented in the sequences. During an infection (counts of uninfected *Karlodinium* cells along with differing infection stages and free-swimming dinospores provided in figure) *Amoebophrya* sequences are more abundant.

CONCLUSIONS AND FUTURE DIRECTIONS

Throughout this dissertation, I explore different types of microbial interactions using the same basic approach: taking multiple samples over time at the same site, using high-throughput sequencing to describe the community, and validating some of the findings using other types of data or experiments. All three types of interactions that I explore (grazing, competition, parasitism) have great ecological relevance in our ecosystem, and teasing apart interactions is impossible without having a temporal dataset. In each case, I find that I am able to take information from the amplicon sequencing dataset and use it to drive additional findings; for example, in Chapter 1, after identifying two putative mixotrophic grazers (*Tetraselmis* sp. and *Gymnodinium sp.*, now *Biecheleriopsis* sp.) based on their co-presence with a flow cytometry population, we were able to demonstrate uptake of *Synechococcus* by both of these. In Chapter 2, we found an unexpected and abrupt switching of sequence variants prior to *Synechococcus* bloom declines that was consistent across multiple *Syenchococcus* clades. We were subsequently able to demonstrate that while two sequence variants were present in one of our cultures, one disappeared under nitrogen-stressed conditions. Finally, in Chapter 3 we were able to track sequences from a known parasite-host culture in the environment throughout 2016 to validate co-occurrence of this pair.

The first and third chapters present a picture of the eukaryotic community at the SIO pier, and we find that there is a "core community" of organisms that are consistently present – including the photosynthetic picoeukaryote *Micromonas*, several cryptophytes and several dinoflagellates like *Heterocapsa*. But most members of the community are not detected all throughout the year. The latter two chapters present data from the year 2016, where there were several large fluctuations in abundance, particularly in the summer, of dinoflagellates, raphidophytes, and the cyanobacteria *Synechococcus*. While this year was likely anomalous in

some ways due to the preceding El Niño conditions, the eukaryotic community had a clearly seasonal signature and temperature was likely an important driver of both eukaryotic and *Synechococcus* dynamics. A *Synechococcus* clade that is not typically abundant at the SIO pier dominated an August 2016 bloom, while eukaryotic abundance (based on flow cytometry-adjusted sequence reads) was much lower in August than earlier in the summer. It certainly makes sense for there to be a succession of blooms throughout the summer; as conditions shift to favor one phytoplankton species, it becomes a target of both 'traditional' predators (like protistan grazers) and parasites. Eventually its ecological dominance wanes and another phytoplankton group might be the most favored.

Dynamics like these are occurring in the thousands (and more) when it comes to a microbial community. While our sequencing results have been able to help describe individual patterns and overall diversity, using this data to actually identify and interpret interactions can be more complicated. Numerous methods have been proposed to identify interacting groups or pairs from large datasets, but ultimately these results are launching points. For every interesting putative interaction or phenomenon, there is great potential for further study, both observational and experimental.

The great diversity found within the genus *Synechococcus* has intrigued researchers for a long time, particularly the co-occurrence of closely related *Synechococcus* variants. Our study has revealed that hidden underneath the broader observations of *Synechococcus* cell abundance are separate dynamics of different clades and different ASVs. Furthermore, though we do see great microdiversity, we suspect that some of the sequence variation might represent a physiological response. We have presented some preliminary evidence for this but the next step will be additional sequencing of cultured *Synechococcus* representatives under nutrient stress.

We found a difference in the sequence variants present between a control culture and a nitrogen-limited culture at a single time point, but sequencing preceding and subsequent time points will allow us to view when the sequence variation shifts. We could also see whether adding back nitrate results in both variants appearing again. This can be repeated with other nutrients as well as other *Synechococcus* clades to see if it is a consistently reproducible phenomenon. If so, it opens up numerous directions to investigate the actual mechanism at play.

Our current understanding of Syndiniales members in the oceans is nascent; few species have been isolated or characterized, and the majority belong to *Amoebophrya* sp. Entire well-supported clades of Syndiniales have never been observed. Using our dataset, we identified some strong correlations between Syndiniales ASVs and other ASVs and propose that those between group II and dinoflagellates might be particularly strong candidates for further investigation. We also found a few causal interactions that include the poorly-understood Syndiniales group III. These could be further probed by using qPCR to track putative parasites and hosts at higher frequency.

The two primary objectives of future work on Syndiniales-host interactions should be (1) validating actual parasite-host pairs through observational or experimental work and (2) better quantifying the dynamics of infections over time by accounting for the different life stages of Syndiniales. I conducted some pilot tests of identifying Syndiniales dinospores or infections using flow cytometry and found that I was able to make gates to count them in culture but the signal was lost when running environmental samples. A more promising direction might be applying a Syndiniales-specific fluorescent probe to seawater samples and then counting with flow cytometry. This would be a very useful direction to pursue because flow cytometry would allow for quantification and for distinction between dinospores and infected hosts (based on cell

size) – this method would likely also be able to distinguish between early, middle, and late stage infections. It would even be possible to flow-sort large cells that have a Syndiniales probe bound and then sequence these cells to identify the associated host. This would be the ideal validation and once a pair is identified it can then be tracked over time in the sequence data. Alternatively, we have demonstrated size fractionation prior to DNA extraction might be able to distinguish between different life stages of Syndiniales. High-throughput sequencing at high temporal resolution has the potential to allow us to track parasite-host pairs over time at the pier, but to then understand infection prevalence, intensity, and mortality rates, we need to be able to discriminate whether sequence reads represent free-living dinospores or actual infected hosts.

One exciting aspect of the data gathered in my dissertation is that it will keep providing a wealth of information moving forward. Questions about any type of eukaryotic interaction can be explored by returning to the sequences from 2016. As new interactions are identified, they can be tracked throughout 2016 as long as the sequences of involved members are known. Similarly, as better analysis methods are developed, they can be applied to this dataset and may yield additional potential directions for further study.