POPULATION DYNAMICS AND DIVERSITY OF *SYNECHOCOCCUS* ON THE NEW ENGLAND SHELF

By

Kristen Rachell Hunter-Cevera

B.S., West Virginia University, 2008
M.B.A., West Virginia University, 2009

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

September 2014

© Kristen R. Hunter-Cevera. All rights reserved.

Signature of Author . . . .
Joint Program in Oceanography / Applied Ocean Science and Engineering
Massachusetts Institute of Technology
and Woods Hole Oceanographic Institution
September 2, 2014

Certified by . . . .
Dr. Heidi M. Sosik
Thesis Supervisor

Certified by . . . .
Dr. Michael G. Neubert
Thesis Supervisor

Accepted By . . . .
Professor Martin F. Polz
Chair, Joint Committee for Biological Oceanography
Massachusetts Institute of Technology and Woods Hole Oceanographic Institution

POPULATION DYNAMICS AND DIVERSITY OF *SYNECHOCOCCUS* ON THE NEW ENGLAND SHELF

By Kristen Rachell Hunter-Cevera

Submitted to the MIT/WHOI Joint Program in Oceanography/Applied Ocean Science and Engineering on September 2, 2014 in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biological Oceanography

ABSTRACT

Synechococcus is a ubiquitous marine primary producer. Our understanding of the factors that determine its abundance has been limited by available observational tools, which have not been able to resolve population dynamics at timescales that match response times of cells (hours-days). Development of an automated flow cytometer (FlowCytobot) has enabled hourly observation of *Synechococcus* at the Martha's Vineyard Coastal Observatory (MVCO) since 2003. In order to ascribe changes in cell abundances to either growth or loss processes, information on division rate is needed. I refined a matrix population model that relates diel changes in the distribution of cell volume to division rate and demonstrated that it provides accurate estimates of daily division rate for both cultured and natural populations. Application of the model to the 11-year MVCO time series reveals that division rate is temperature limited during winter and spring, but light limited during fall. Inferred loss rates closely follow division rate in magnitude over the entire seasonal cycle, suggesting that losses are mainly generated by biological processes. While *Synechococcus* cell abundance, division rate, and loss rate demonstrate striking seasonal patterns, there are also significant shorter timescale variations and important multi-year trends that may be linked to climate. Interpretation of population dynamic patterns is complicated by the diversity found within marine *Synechococcus*, which is partitioned into 20 genetically distinct clades. Each clade may represent an ecotype, with a distinct ecological niche. To understand how diversity may affect population dynamics, I assessed the diversity at MVCO over annual cycles with culture-independent and dependent approaches. The population at MVCO is diverse, but dominated by clade I representatives throughout the year. Other clades were only found during summer and fall. High through-put sequencing of a diversity marker allowed a more quantitative investigation into these patterns. Five main *Synechococcus* oligotypes that comprise the population showed seasonal abundance patterns: peaking either during the spring bloom or during late summer and fall. This pattern strongly suggests that features of seasonal abundance are affected by the underlying diversity structure. *Synechococcus* abundance patterns result from a complex interplay among seasonal environmental changes, diversity, and biological losses.

Thesis Supervisors:
Dr. Heidi M. Sosik
Title: Senior Scientist, Woods Hole Oceanographic Institution
Dr. Michael G. Neubert
Title: Associate Scientist with Tenure, Woods Hole Oceanographic Institution

Acknowledgments

This has been an incredible journey, both scientifically and personally. The people thanked below are only a few who made it that.

To my advisors, Mike and Heidi, I cannot express how grateful I am for your guidance, mentoring, and support for the past 5 years. I am honored to be your student and have learned so much from both of you. I will miss our Friday noon meetings.

To the rest of my committee: Anton thank you for opening up your lab to me. I'm incredibly grateful to have spent time there and thank you for your guidance and support.

Andy, thank you for your expert statistical knowledge and advice, positive and entertaining comments, and being willing to follow any rabbit hole for the model.

Penny, thank you for your wise perspective, excellent advice, and borrowing of your lab instruments.

To John Waterbury, thank you for teaching me the isolation techniques and letting spend time in your lab. I could not have gotten the culture isolations done without your help.

To Rob and Alexi, thank you for all your help and advice on this project over the years, notwithstanding the maintenance and support of FCB. I am very grateful to you both.

To Becky, thank you for serving as chair for both my proposal and thesis defense, and for always letting me use the FACS.

To Sam for good science and life advice and instrument help along the way.

To past and current members of the Sosik Lab, in particular Emily, Emily and Taylor - thank you for all your help and support over the years.

To Kasia Hammer, thank you for all the help in lab, especially with the sample pro-

cessing. I am truly grateful.

To the other past and present members of the Post lab: thank you Erin, Heather and Tom.

To members of the Bay Paul Center, thank you for all the help and being willing to answer my non-stop questions. I feel grateful to have been able to spend time with you.

To the math ecology group, thank you for the always stimulating discussions and help, and in particular to Hal Caswell, for always being willing to see where an idea will lead.

To Freddy Valois and Athena Aicher, thank you for the culturing advice and assistance with the cultures.

To my cohort, Amalia, Britta, Liz, Jill and Santiago, for your friendship, support and beach breaks. Grad school would have been very different without you.

To Facilities and Security, thank you for making things run and for checking in on me during late nights.

To my cultures, for growing away from the ocean.

To members of APO, particularly Julia, Jim, Meg, Lea, Tricia and Marsha. We are lucky as students to have you.

To Dehann, for your love and support and making the last year unforgettable. I'm so glad I went on the boat.

To my family, for all their love, support and encouragement since I was young. To my dad, for always being proud of me, to my brother for keeping me humble, and to my mom for guidance, mentoring and help with the cultures. I would not be writing this sentence without you.

Contents

List of Figures

13

List of Tables

Chapter 1

Introduction

What determines the abundance of an organism? What are the factors that allow it to grow, reproduce and survive? These are especially important questions for the marine cyanobacterium *Synechococcus*. This picophytoplankter is present at high cell concentrations across the world's oceans and is an important contributor to primary production [73, 59], fixing up to 20% of carbon in coastal waters [59]. Since its discovery in 1979 [46, 113], many studies have focused on answering these questions. For excellent reviews and key papers of the physiology and ecology of *Synechococcus* the reader is referred to [114, 92, 91].

A complete answer to these questions requires both knowledge of the physiology of the organism and the environmental conditions under which it lives. How these two factors interact will determine abundance. In particular, the population abundance at any one time will be determined by how fast new cells are being produced (rate of cell division) and how fast they are begin removed from the system (loss rate). The rate of cell division is determined by the physiological state of the cell. Division rate can be limited by different environmental variables. Temperature can limit the rate at which metabolic activities occur. Light can limit the rate at which carbon is fixed and ATP is produced. A short supply of nutrients will limit the materials needed for new cell material construction (macro nutrients) or cofactors needed for enzymes and reaction centers (micro nutrients). The rate of cell division therefore provides information about possible physiological limitations the cell is experiencing.[1] Limitation by an environmental resource is termed 'bottom-up' control on cell abundance.

For marine bacteria, cells can be removed quickly from the system via predation by heterotrophic microzooplankton (including heterotrophic nanoflagellates, ciliates, di-

---

[1]Growth rate is also used as a term for division rate. For clarity, only the term division rate will be used to avoid confusion with the term net growth rate, discussed in a later chapter.

noflagellates and others [52]) or via viral lysis [63]. Knowledge of predation or infection pressure informs our understanding of the biological environment that *Synechococcus* encounters. Control of cell abundance by loss processes is termed 'top-down' control. These grazing organisms and viruses are a part of what is called the 'microbial loop': a trophic pathway through which bacteria enter the marine food web. It is important to understand how *Synechococcus* is lost from marine systems. The carbon fixed by *Synechococcus* will have different fates if ingested by a predator or released into the environment by viral lysis. If ingested, this carbon has the potential to work its way through the food chain; if lysed, heterotrophic bacteria will most likely remineralize the carbon. Predators and viruses are not the only biological elements with which *Synechococcus* can interact. Other heterotrophic bacteria have the potential to exert 'sideways' control on the population. For the sister genus *Prochlorococcus*, division rates of cultures depend on which heterotrophic bacteria are grown in co-culture [95]. Morris et al. [70] hypothesized that some of the potential benefits to *Prochlorococcus* were removal of harmful oxygen species produced during photosynthesis, but other benefits could be increasing carbon dioxide concentration, or removal of waste products [95]. It is likely that similar relationships exist for *Synechococcus*. Anecdotal observations of *Synechococcus* in culture suggest that non-axenic cultures are somewhat more stable and more robust than axenic cultures (J. Waterbury, personal communication).

Understanding the interactions between *Synechococcus* and environmental and biological factors is further complicated by the fact that *Synechococcus* is a genetically and physiologically diverse genus [92]. Different strains demonstrate different physiological responses to environmental factors, as well as interactions with heterotrophic grazers [80, 68, 6, 83]. Knowledge of how *Synechococcus* diversity changes over time is important for understanding how the population as whole responds. Of equal interest is how such 'microdiversity' is maintained among such closely related organisms. This is a rather intriguing case of the 'Paradox of the Plankton' posed by Hutchinson in 1961 [41]: how can so many different phytoplankton species simultaneously coexist when they appear to be all competing for the same limited resources? Maintenance of diversity within the phytoplankton is a question that is still being explored today. One of the resolutions relies on changing environmental conditions. If the timespan of environmental conditions that support the growth of one type of phytoplankton over another is shorter than the time it would take for competitive exclusion then both phytoplankton types can persist in the environment [41]. Environmental instability and heterogeneity are thought to be major mechanisms that allow maintenance of

diversity [89], and, indeed, the marine environment is dynamic. The regular seasonal cycle, as well as short term variability in weather patterns, wind, mixing, and advection will continuously affect the light, temperature, and nutrient environment that a cell experiences. Additional mechanisms thought to maintain diversity are selective interactions with predators and viruses [89]. It is thought that each clade (or group of clades) within the *Synechococcus* genus has a distinct environmental niche [3] that permits persistence of diversity within these cyanobacteria.

How all of these possible factors combine to produce a change in *Synechococcus* abundance over time is a complicated and complex problem. Given that the interactions happen on time scales of seconds to days, investigation is difficult with conventional sampling and observational tools. To fully understand the system, measurements and observations need to be made at the appropriate temporal resolution. Current advances in automated technology have begun to address this issue. The development of an automated, submersible flow cytometer (FlowCytobot, [74]) and other automated imaging flow cytometers and systems (Imaging FlowCytobot [75], FlowCam, [96]) have enabled an unprecedented view into the dynamics of marine phytoplankton. Analysis of this highly resolved data confirms the complexity of the system. For *Synechococcus*, daily resolution for extended periods has illustrated dynamic increases and decreases over the course of an annual cycle and highlights the complicated problem of trying to unravel these patterns.

Observations of cell concentration with high temporal resolution are critical for understanding changes in cell abundance. However, observations of cell concentration alone can be difficult to interpret. Estimates of either division rate or loss rate are needed to separate the effects of growth and loss processes on the change in abundance. For example, if there was no change in cell abundance from one day to the next, one explanation could be that there was little cell division and little predation. However, an alternative scenario, that would produce the same results, is high cell division coupled with high losses. It is important to be able to distinguish between these two scenarios as they have different ramifications for carbon and nutrient cycling and also for our understanding system dynamics.

The overarching goal of my thesis is to understand the population dynamics of *Synechococcus* and how those dynamics depend on the underlying diversity structure. In order to address these subjects, I have undertaken several research activities at the study site of Martha's Vineyard Coastal Observatory (MVCO), located on the New England Shelf. In Chapter 2, I refine and validate a matrix population model that when fit to time series of cell size distributions is able to estimate in situ daily

division rates. Estimates of division rate on the same timescale at which we can measure cell concentration reveals underlying causes of seasonal population dynamics. In Chapter 3, I apply the matrix population model to an 11-year time series of cell size distributions obtained from FCB since 2003 at the MVCO. Analysis of daily division rates and cell concentration provides insights into the environmental factors that produce the repeatable seasonal patterns, as well as causes of variation in those patterns. In Chapter 4, I explore the potential diversity of *Synechococcus* present at MVCO over annual cycles. Culture isolations, analysis of clone library data, and flow cytometry reveal that the *Synechococcus* assemblage is diverse and hints at possible seasonal diversity patterns. To further explore these patterns and how they may relate to the overall population dynamics, I carried out quantitative estimation of different *Synechococcus* oligotypes. I report on these analyses in Chapter 5, and they reveal that each oligotype exhibits a distinct annual phenology. Features of the seasonal cycle of *Synechococcus* abundance are likely to result from the presence of different oligotypes. I summarize these findings, draw overall conclusions, and suggest future research endeavors in Chapter 6.

Chapter 2

Diel size distributions reveal seasonal growth dynamics of a coastal phytoplankter

2.1 Introduction

Marine phytoplankton contribute approximately 50% of global net primary production [27], mediate global biogeochemical cycles, and form the base of marine food webs. It is vital that we understand the factors that govern their abundance, the more so in light of on-going climate change. Key to this is an understanding of the rate at which phytoplankton cells divide under different environmental conditions.

Division rate cannot be measured from changes in cell abundance alone, as changes in abundance result from interactions between cell division and other processes such as predation, advection, sinking, and mixing. Further, we lack approaches that can resolve these processes on scales relevant to the cells' responses to their environment. To overcome this, estimates of abundance and division rate are needed on time scales of hours to days and extending for weeks, months, and ultimately years. While some progress has been made with automated sampling [14], a practical method for estimating division rates across this wide range of scales has remained elusive. Conventional methods require incubations [54, 118] or sample manipulation and handling [65, 13, 10], neither of which can be feasibly conducted at daily resolution for extended duration.

For the important class of picophytoplankton ($\leq 2\mu$m in diameter), estimation of division rates has been attempted from in situ diel changes in cell size. During

daylight, individual cells photosynthesize and increase in volume. The increase in cell volume during daylight hours provides a minimum estimate of the daily division rate [25]. This approach has been used to study *Synechococcus*, *Prochlorococcus*, and picoeukaryotes in the open ocean [25, 11, 111]. A major limitation of this approach is its implicit assumption that (at the population level) cell growth and division are separated in time. While most cell division occurs around dusk [114, 111], in fact these processes have been observed to occur simultaneously throughout the day in cultures of *Synechococcus*, especially when division rate is high ($> 0.7 \, \mathrm{d}^{-1}$) [10, 114, 42]. Under such conditions, this approach underestimates division rate. Sosik et al. [98] proposed an alternative method based on a matrix population model that represents changes in cell sizes and allows for simultaneous growth and division. This model can be fit to time series of cell size distributions and the fitted model provides an estimate of the daily division rate. A key advantage both approaches share is that they do not depend on cell concentration. This is especially important in dynamic coastal systems where complex interactions between physical and biological processes can produce patchy plankton distributions [100].

While the approach of Sosik et al. [98] may be powerful, its efficacy and accuracy have never been evaluated. Here we provide the first direct test of the model-based approach. We show that, for both cultured and natural *Synechococcus*, the model-based approach is, on average, in excellent agreement with methods based on cell counting. We apply the approach to observations collected by an automated submersible flow cytometer (FlowCytobot; FCB) [74] over an annual cycle at the Martha's Vineyard Coastal Observatory (MVCO). Our analysis reveals, with unprecedented resolution, a distinct seasonality of division rates. It also establishes that division and loss processes are tightly coupled throughout the year. Despite this coupling, dramatic seasonal changes in cell abundance occur because small but systematic differences favor net growth (or loss) for extended periods (weeks to months).

2.2 Model Construction

We estimate the division rate with a matrix population model[1] (based on [98]). In this section, we describe the model and its fitting. We make two assumptions. First, within a single day, cell growth is determined by incident radiation, with temperature, nutrient availability, and other factors operating at longer time scales. Second, the

---

[1]The formulation and analysis of matrix population models are extensively covered in Caswell [16]

odds of a cell dividing depend only on its size and are constant for cells within a discrete size class. With descriptions of how cell growth depends on light and cell division depends on size, the model predicts the cell size distributions over the course of a day.

We begin by dividing the cells into m size classes. The cells in class i have volumes between ν_i and ν_{i+1} where the class boundaries are equally spaced on a logarithmic scale,

$$\log_2 \nu_i = \log_2 \nu_{\min} + (i-1)\Delta_\nu, \ \text{for } i = 1, 2, \cdots, m. \tag{2.1}$$

$\nu_{\min}$ is the minimum cell volume and Δ_ν is the class width. Here we take $\nu_{min} = 2 \cdot 10^{-5} \mu\text{m}^3$, $m = 57$, and $\Delta_\nu = 0.125$. Let $\mathbf{n}(t)$ be the m-by-1 vector whose elements, $n_i(t)$, are the number of cells in class i at time t. The population vector at time $t + dt$ is given by:

$$\mathbf{n}(t + dt) = \mathbf{A}(t; \boldsymbol{\theta})\, \mathbf{n}(t). \tag{2.2}$$

The element $a_{ij}(t; \boldsymbol{\theta})$ of the m-by-m projection matrix $\mathbf{A}(t; \boldsymbol{\theta})$ is the number of cells in class i at time $t + dt$ per cell in class j at time t. Each of these elements is a function of time and depends on parameters collected in $\boldsymbol{\theta}$. Given these parameters and an initial condition $\mathbf{n}(0)$, the model projects the population distribution throughout the day. The daily division rate is

$$\mu = \ln\left(\frac{N(24)}{N(0)}\right), \tag{2.3}$$

where $N(0)$ and $N(24)$ are the total cell numbers from the model at the beginning and end of the day. As summarized in Fig. S1, the model allows a cell to undergo only one of two transitions in a single time step: division or growth. (Cells that neither divide nor grow survive in the same size class.) The time step dt should be sufficiently small such that this is a reasonable assumption; we set $dt = 10$ min.

2.2.1 Division

We assume that a fraction $\delta(t, \nu_j; \boldsymbol{\theta})$ of the cells in size class j divide in half during one time step, and that the odds of dividing depend on cell size and time of day. In particular, we set

$$\delta(t, \nu_j; \boldsymbol{\theta}) = \begin{cases} 0, & \text{for } t \leq t^*, \\ \left(\frac{\nu_j^b}{1+\nu_j^b}\right) \delta_{max}, & \text{for } t^* < t < 24, \end{cases} \tag{2.4}$$

with $b > 0$. Cells do not divide before time t^*, an assumption that is supported by observations of natural and cultured *Synechococcus* [10, 114], including the culture used in this study. We take $t^* = 6h$. After t^*, the division probability δ is an increasing function of cell size. δ_{max} is the maximum fraction of cells that divide and the parameter b determines the shape of δ (Fig. 2-2A).

We assume when a cell divides, each daughter cell is half the size of the original. If division would produce cells that are smaller than $\nu_{\min}$, those daughters are instead assigned to the first size class. This occurs for all mother cells in classes less than or equal to k,

$$k = 1 + \frac{1}{\Delta_\nu}. \tag{2.5}$$

Note that Δ_ν must be chosen such that $1/\Delta_\nu$ is an integer. Division is therefore described by elements along the first row of $\mathbf{A}(t; \boldsymbol{\theta})$:

$$a_{1,j}(t; \boldsymbol{\theta}) = 2\delta(t, \nu_j; \boldsymbol{\theta}), \text{ for } j = 1, 2, ...k, \tag{2.6}$$

and by the supradiagonal

$$a_{j+1-k,j}(t; \boldsymbol{\theta}) = 2\delta(t, \nu_j; \boldsymbol{\theta}), \text{ for } j = k+1, ...m. \tag{2.7}$$

2.2.2 Growth

Of the cells that do not divide, a fraction $\gamma(t; \boldsymbol{\theta})$ grow into the next largest size class. This fraction is independent of cell size, but depends upon incident radiation $E(t)$. We assume that $\gamma(t; \boldsymbol{\theta})$ is a piecewise linear function of $E(t)$ (Fig. 2-2B):

$$\gamma(t; \boldsymbol{\theta}) = \begin{cases} \gamma_{\max} (E(t)/E^*) & \text{for } E(t) < E^*, \\ \gamma_{\max}, & \text{otherwise.} \end{cases} \tag{2.8}$$

When $E(t) \geq E^*$, the fraction of cells that grow into the next size class is γ_{max}. Cells in the largest size class do not grow. The elements of $\mathbf{A}(t; \boldsymbol{\theta})$ that correspond to cell growth occur along the first subdiagonal:

$$a_{j,j+1}(t; \boldsymbol{\theta}) = \gamma(t; \boldsymbol{\theta}) \left[1 - \delta(t, \nu_j; \boldsymbol{\theta})\right], \text{ for } j = 1, 2, \cdots m - 1. \tag{2.9}$$

2.2.3 Stasis

Cells that neither divide nor grow remain in the same size class. These transitions appear on the main diagonal of $\mathbf{A}(t; \boldsymbol{\theta})$:

$$
a_{j,j}(t, \boldsymbol{\theta}) = \begin{cases} [1 - \gamma(t, \boldsymbol{\theta})] \, [1 - \delta(t, \nu_j; \boldsymbol{\theta}))] + 2\delta(t, \nu_j; \boldsymbol{\theta}), & \text{for } j = 1 \\ [1 - \gamma(t, \boldsymbol{\theta})] \, [1 - \delta(t, \nu_j; \boldsymbol{\theta})] \,, & \text{for } 2 \leq j \leq m - 1 \\ [1 - \delta(t, \nu_j; \boldsymbol{\theta})] \,, & \text{for } j = m \end{cases} \quad (2.10)
$$

All matrix elements not assigned by equations (2.6)–(2.10) are zero.

2.2.4 Subpopulations

When the initial distribution is unimodal, the model preserves this feature over the course of the entire day. In the majority of our laboratory experiments, however, bimodal size distributions developed and then disappeared during the day (Fig. 2-3). This phenomenon also occurs in natural assemblages at MVCO, although more subtly and less frequently. Bimodal size distributions in the laboratory experiments typically first appeared during the start of cell concentration increase and disappeared a few hours before cell concentration stopped increasing for the day. This timing and the magnitude of increase in volume apparent for larger cells (see Fig. 2-3B) suggests that the underlying cause was early rounds of cell division [10] with a portion of these newly divided cells staying attached for much longer than typical in the field. Other possibilities include a portion of the culture that exhibited different cell size dynamics (such as growing much faster or having a different diel timing for cell division). This is a possibility as even in clonal culture, phenotypic differences have been observed after spontaneous mutations over many generations [119, 90].

We can accommodate bimodal size distributions by including in our model two subpopulations, $\mathbf{n}_1(t)$ and $\mathbf{n}_2(t)$, each governed by its own projection matrix and parameters. The total number of cells at time t is given by $N(t) = N_1(t) + N_2(t)$ where $N_1(t)$ and $N_2(t)$ are the totals for the two subpopulations. The modeled overall daily growth rate is found from Eqn. 2.3.

To complete the model, it is necessary to specify the initial size distributions of the two components. The observed initial size distributions are consistent with a mixture of 2 lognormal components with different mean parameters $\bar{\nu}_1$ and $\bar{\nu}_2$, common variance parameter σ^2, and mixing proportion ψ (defined as the proportion of cells in the first component).

We therefore specify the i-th element of $\mathbf{n}_\ell(t^*)$ as

$$\frac{\psi_\ell}{c\sqrt{2\pi}\left[\frac{1}{\Delta_\nu}\log_2(\frac{\sigma}{\nu_{min}})+1\right]}\exp\left\{\frac{-\left[\frac{1}{\Delta_\nu}\log_2(\nu_i/\bar\nu_\ell)\right]^2}{2\left[\frac{1}{\Delta_\nu}\log_2(\frac{\sigma}{\nu_{min}})+1\right]^2}\right\} \qquad (2.11)$$

for $\ell = 1, 2$. We choose the constant c so that $N_\ell(t^*) = 1$. The parameters ψ_1 and ψ_2 specify the fractions of the total population that are in the first and second subpopulations. As such, $\psi_1 = \psi$ and $\psi_2 = 1 - \psi$.

The model outlined above contains a total of 12 parameters. For each subpopulation, the parameters b, δ_{max}, γ_{max}, and E^* specify the projection matrices. In addition to these eight, we have four additional parameters that describe the starting populations: ψ, $\bar\nu_1$, $\bar\nu_2$, and σ^2. We assume that all of these parameters are constant within a day, but may change between days. The variables, parameters and constants that are needed to completely specify the model are summarized in Table 2.1.

Parameter estimates do not depend on the concentration of cells, but only on the proportion of cells in each size class. Importantly, we find that our two-subpopulation model is able to reproduce the range of cell size distribution patterns that typically occur in the data (Fig. 2-4). We also find that by eliminating the first 6 hours of the day from the model fitting results in better representation of the observed cell size distributions. This suggests that there is a feature of the *Synechococcus* growth and division cycle that we are not capturing in our equations with regard to cell size dynamics right after dawn. However, this does not appear to impact the model's ability to estimate an accurate division rate.

2.3 Parameter Estimation

To estimate the parameters, we fit the model to hourly observations of the number of cells in each size class. It would be natural to assume that our observations have a multinomial distribution. Our data, however, are over-dispersed relative to the multinomial. To account for this overdispersion in the data we instead assume that the observation at hour t has a Dirichlet-multinomial distribution [45] whose probability density function is given by

$$f(\hat{\mathbf{n}}(t);\mathbf{w}(t,\boldsymbol\theta)) = \frac{\hat{N}(t)!}{\hat{n}_1(t)!\,\hat{n}_2(t)!\,\cdots\,\hat{n}_m(t)!}\left[\frac{\Gamma(s)}{\Gamma(\hat{N}(t)+s)}\right]\frac{\prod_{i=1}^m\Gamma(\hat{n}_i(t)+sw_i(t,\boldsymbol\theta))}{\prod_{i=1}^m\Gamma(sw_i(t,\boldsymbol\theta))}. \tag{2.12}$$

In Eqn. 2.12, Γ is the gamma function, $\hat{n}_i(t)$ is the number of cells observed in size class i at time t, $\hat{N}(t)$ is the total number of cells observed at time t ($\hat{N}(t) = \sum_{i=1}^{m} \hat{n}_i(t)$) and $\mathbf{w}(t)$ is the distribution of cells in each size class obtained from the model:

$$\mathbf{w}(t) = \frac{\mathbf{n}_1(t) + \mathbf{n}_2(t)}{N(t)}. \tag{2.13}$$

The parameter s is a precision parameter; the larger s the less the dispersion and the closer the density (Eqn. 2.12) is to the multinomial density. To use the Dirichlet-multinomial distribution, we must specify this parameter, which brings the total number of parameters to 13 with s now included in $\boldsymbol{\theta}$.

Our estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$, maximizes the likelihood function:

$$L(\boldsymbol{\theta}|\hat{\mathbf{n}}) = \prod_{t=s}^{24} f(\hat{\mathbf{n}}(t); \mathbf{w}(t, \boldsymbol{\theta})) \tag{2.14}$$

subject to the constraints listed in Table 2.1.

2.3.1 Confidence Interval Construction

The confidence intervals reflect the uncertainty surrounding the model estimate due to sampling error, given the assumption that the underlying model structure is a correct representation of cell dynamics. One can construct a confidence interval for any parameter in $\hat{\boldsymbol{\theta}}$ with the Fisher information matrix (FIM), $\mathcal{I}(\hat{\boldsymbol{\theta}})$, the entries of which are:

$$-\mathrm{E}\left[\frac{\partial^2 \log L(\hat{\boldsymbol{\theta}})}{\partial \theta_i \partial \theta_j}\right], \tag{2.15}$$

were E denotes the expectation. Here we calculate the *observed* FIM as the $n \times n$ matrix $\mathbf{I}(\hat{\boldsymbol{\theta}})$, with elements

$$-\frac{\partial^2 \log L(\hat{\boldsymbol{\theta}})}{\partial \theta_i \partial \theta_j}. \tag{2.16}$$

We use the asymptotic normality of $\hat{\boldsymbol{\theta}}$ [7] to construct confidence intervals around a particular parameter in $\hat{\boldsymbol{\theta}}$ as

$$\theta(j) \pm C\left(\mathbf{I}_{jj}^{-1}(\hat{\boldsymbol{\theta}})\right)^{\frac{1}{2}}, \tag{2.17}$$

where $\mathbf{I}_{jj}^{-1}(\hat{\boldsymbol{\theta}})$ refers to the jth diagonal entry of the inverse observed FIM and C is the desired critical value of the normal distribution (i.e., 1.96 for 95% confidence interval). Second derivatives were estimated with finite difference calculations (see Appendix). For the diagonal elements of the FIM we used the center difference rule when the maximum likelihood estimate of the parameter was away from the bounds and a forward difference or backward difference equation when the parameter was on the bounds. In the case of mixed partial derivatives for the rest of the elements in the FIM, we used a combination of forward, backward, and center difference equations to calculate the second derivative depending on whether the parameter was on the bounds or not.

We use this approach to obtain confidence intervals for the division rate by treating the division rate as a parameter. Instead of calculating it from the other parameters, we estimate it directly and calculate δ_{max} of one subpopulation instead. The relationship between δ_{max} of one of the subpopulations and the calculated division rate is monotonically increasing. Therefore, if we specify a division rate, we should be able to solve for the corresponding δ_{max}, with all other parameters held constant. We use a root finding solver (fzero solver offered in MATLAB) to identify the δ_{max} that produces a specified division rate. We then run the model forward with the found δ_{max}. Given some parameter combinations, certain division rate values are not feasible within the potential range of δ_{max}, [0,1]. This has the potential to cause issues when calculating the second derivatives for the observed FIM. Generally, however, we do not run into this problem often. Evaluating the FIM when parameters are at the bounds may violate some of the asymptotic normal assumptions that allow us to calculate confidence intervals for parameters in $\hat{\boldsymbol{\theta}}$. We investigated this by simulating data with parameters chosen on the bounds (with sampling from the Dirichlet multinomial distribution), fitting this simulated data with our model, and calculating the confidence intervals. We found that a majority of the resulting confidence intervals contained the true division rates, and thus feel confident in this approach even if parameter values were on the bounds.

2.3.2 Model Differences from Sosik et al. (2003)

The model differs from that presented in Sosik et al (2003) in three ways. First, two subpopulations are allowed to exist and behave according to their own growth (γ) and division (δ) functions. Second, the division function used here is different from that presented in Sosik et al. (2003):

$$\delta(t, \nu_j; \boldsymbol{\theta}) = \begin{cases} 0, & \text{for } t \leq 6, \\ \left(\frac{a\nu_j^b}{1+a\nu_j^b} \right) \delta_{max}, & \text{for } t > 6, \end{cases} \qquad (2.18)$$

as our function no longer contains the a parameter. Estimated parameters of simulated multinomial sampled data demonstrated an inverse relationship between parameters δ_{max} and a, indicating an unnecessary flexibility for the division function. Division rate estimates for simulated data were nearly identical between the two division function formulations. The third difference between the models is that a starting distribution is fit according to a mixture of log normal distributions for the two subpopulations. In Sosik et al. [98], the observed distribution at hour 0 was used as the starting distribution.

Parameter estimation is also different from Sosik et al. [98], who used a nonlinear least squares approach. Here, we use a maximum likelihood approach. We also only fit the model to a partial day. These changes from the original model version are supported (and some initially inspired) by our evaluation of the model's ability to estimate division rate of both cultured and natural *Synechococcus* populations.

2.4 Materials and Methods

2.4.1 Culture Setup and Sampling

To evaluate the model's ability to estimate division rates, we applied it to daily cell size distributions of an MVCO *Synechococcus* isolate grown under a range of temperature and light conditions. The *Synechococcus* strain used for this study was isolated from coastal surface waters at MVCO in May of 2006. While this isolate did not go through any clonal isolation steps (i.e., no dilution to extinction, sorting, or plating), sequencing of the diversity marker *ntcA* for this culture shows only closely related representatives that belong to clade I of marine *Synechococcus* (see below). The isolate was grown as batch cultures in 1-L jacketed vessels with SNAX medium [114] with trace metal amount reduced to 20% to minimize precipitation. Cultures were kept in exponential growth at cell concentrations spanning the range typical at MVCO. A range of division rates was achieved by varying temperature and light intensity. Temperature was controlled between 10 °C and 28 °C by pumping water through the vessel jacket approximately 1 L min^{-1} from a water bath. Vessels were illuminated by two 32 W white fluorescent lamps with controllable light output and a 14:10 hour light:dark cycle. Light intensity was controlled and recorded by a MiniLab

USB data acquisition interface connected to a scalar irradiance meter (Biospherical Instruments, QSL 100). Incident light followed a sinusoidal pattern with maximum intensity 80 μmol quanta m^{-2} s^{-1} inside the vessels. Black plastic screens were used to decrease light intensity to 65%, 42%, 27% and 18% (1 to 4 layers of mesh, respectively) of maximum. To reduce the tendency of cells to clump and stick to the walls of the vessel, the walls were siliconized and cultures were mixed at about 200 rpm by glass paddles.

Cultures were sampled with a laboratory version of FlowCytobot [74]. Data analysis and enumeration of *Synechococcus* cells were as described in [98]. FCB measures side angle scattering, which is converted to cell volume [74]. The model was applied to cell size distributions from individual days of batch grow outs (Fig. 2-5). For estimate comparison, division rates from the culture were calculated with Eqn. 2.3 using cell concentrations at dawn of one day and dawn of the next day. Culture observations were included in the final data set for model comparison only if they exhibited (1) division patterns phased to the diel cycle and (2) division rate consistent with other days under similar conditions (i. e., outlier or negative division rates were excluded).

2.4.2   Dilution Series Experiments

To evaluate the model's ability to estimate division rate from natural *Synechococcus* populations and the possible affect of grazers on model estimates, we conducted 12 dilution series experiments [54] (6 days in June 2012 and 6 days in October 2012, Table 2.2) with water from Woods Hole Harbor and sampled by FCB. For each individual experiment, seawater at the farthest point off the Woods Hole Oceanographic Institution Iselin dock was collected 2 hours before dawn via bucket sample. Water was passed through a 233 μm mesh to exclude larger zooplankton predators, but not the protozoa that typically feed on *Synechococcus*. Water was kept in the dark in 24-L carboys until experiments were started within 1.5 hours after sampling. A portion of the water was filtered through 20 μm mesh (by gravity) before filtration by peristaltic pump with a 0.2 μm inline Versapor capsule filter (Pall Corporation) to yield filtered seawater. The inline filter was acid-washed before use and in between every two dilution series experiments by pumping 10% HCl through the filter, followed by flushing with 2 L of Milli-Q (Millipore Corporation) water. Before filtrate was collected, approximately 1 L was passed through the inline filter and discarded. Acid-cleaned 1.25-L polycarbonate bottles were triple rinsed with whole seawater before being filled with portions of whole and filtered seawater. Water was combined

to yield dilutions corresponding to fractions of either [0.1, 0.3, 0.6 and 1] or [0.2, 0.4, 0.6 and 1] of whole seawater. Each dilution level was prepared in triplicate and total final volume in each bottle was 1 L. Lower fractions were used for dilution in summer since *Synechococcus* abundance was 10-fold higher than in autumn. We aimed to have > 200 cells mL^{-1} to resolve hourly cell size distributions with FCB. All bottles were shaded with 2 layers of window screen. Incubations were conducted at ambient water temperature in a flow-through seawater tank where FCB was also located. For each experiment two bottles corresponding to the highest dilution and whole seawater were sampled by FCB. Data processing was the same as that for FCB as described in [98]. Irradiance was measured with the surface portion of a HyperPro radiometer (Satlantic LP). For a few days, when radiation data could not be obtained, radiation data from MVCO (located 15 miles south from Woods Hole Harbor) was used.

Bottles not connected to the FCB were sampled at time 0 (dawn local time), 24 and 48 hours. Samples were preserved with glutaraldehyde to a final concentration of 0.1% and incubated 10 min before being stored in liquid nitrogen until later flow cytometric analysis. *Synechococcus* in preserved samples were enumerated with a FACSCaliber (BD Biosciences) flow cytometer. *Synechococcus* cells were identified on the basis of their characteristic orange fluorescence from the accessory pigment phycoerythrin [73]. Net division rate was calculated with Eqn. 2.3, which assumes exponential growth and loss processes. We fit the data (net division against dilution level) with either a one-phase or two-phase linear regression model as some data suggested nonlinear dilution response [29]. To determine if a linear relationship between net growth rate and dilution factor was appropriate, we tested a one-phase regression model against a two-phase regression model for each dilution series experiment [36]. The one-phase model is the standard linear regression model: $Y = \beta_O + \beta_1 X + \epsilon$, where Y is the observed data, X is the regressor, β_O and β_1 are the intercept and slope parameters and ϵ is the error. The two-phase model is $Y = \beta_O + \beta_1 X + \beta_2 (X - X_C)I(X) + \epsilon$, where X_C is a change point at which the data is better represented by a different linear form and $I(X)$ is an indicator function, such that $I(X) = 1$ if $X \geq X_C$ and 0 otherwise. First,we fit both the one-phase and two-phase model by minimizing the residual sum of squares (RSS). We then calculated an F statistic, which is well approximated by an F distribution with 3 degrees of freedom in the numerator and n-4 degrees of freedom in the denominator [36]:

$$F_{stat} = \frac{\frac{(RSS_1 - RSS_2)}{3}}{\frac{RSS_2}{n-4}} \sim F(3, n-4),$$

and n is the number of data points. We rejected the null hypothesis of a single phase in favor of two phases at significance level $\alpha = 0.05$. In the case where the null hypothesis was not supported, the division rate was taken as $\hat{\beta}_O$ from the fitted two-phase model (y-intercept - net growth rate extrapolated to 0 fraction of whole sea water). Confidence intervals for the intercept were constructed from the profile log likelihood of β_O and the likelihood ratio test. The confidence interval included values of β_O such that the likelihood ratio ($2[\log L(\hat{\beta}_O) - \log L(\beta_O)] \sim \chi_1^2$, from profile log likelihoods) would not be rejected at significant level $\alpha = 0.05$. Grazing rate was calculated as the difference between the intercept rate and the whole seawater net growth rate.

We let each experiment run for a total of two days, treating each day as a separate case to compare model and dilution-based division rate estimates. We recognize that bottle effects may influence dilution series experiments that run for more than 24 hours, but in this case, our goal was not to assess precisely in situ division rate, but rather to compare division rates with those from the model and to investigate the possible impact of grazing on model estimates. In most of the experiments conducted, grazing appeared to increase during the second day, which provided an effective dataset for the latter objective. For each experimental point, division rate was only calculated over a 24-hour period.

Five of the twelve dilution series experiments demonstrated no change in net growth rate across the dilution fractions. Considering the cell concentration at which a linear net growth rate response was observed in the other experiments, these results are consistent with cell concentration of *Synechococcus* being above the grazer ingestion saturation threshold even in the most diluted bottle. Therefore, no division rate estimate could be obtained for these experiments and the data was not used. The remaining seven dilution series experiments provided division estimates that could be compared with the matrix population model (Table 2.2).

2.5 Results

2.5.1 Culture Experiments

To evaluate the model's ability to estimate division rates, we applied it to daily cell size distributions of an MVCO *Synechococcus* isolate grown under a range of temperature and light conditions. For this culture we estimated the division rate with standard counting methods. We found that across a range of conditions the

model-based and count-based estimates were in good agreement on average (Fig. 2-6). The level of overall agreement can be measured by the accuracy component C_b of the concordance correlation coefficient [55]. Specifically, C_b lies between 0 and 1 and measures how far the best fit linear relationship between 2 estimates deviates from the one-to-one line. When $C_b = 1$, the two lines are identical. For the data in Fig. 2-6, the estimate of C_b is 0.989 with a 95% confidence interval of (0.949, 0.997). Such extremely high accuracy demonstrates that, on average, our model works as well as standard counting methods for estimating division rates across a range of growth conditions, diel patterns, and other controlling factors.

We did encounter days when model and cell count division estimates did not agree, in particular, at higher temperatures (water temperatures at MVCO do not exceed 22 °C) and low light. The experiments with these discrepancies tended to occur sequentially (i.e., days in a particular batch experiment) and had an unusual cell size distribution pattern (abrupt increase then decrease of cell volume in the span of a few hours) that the model was not able to reproduce. Other batch experiments grown under these same light and temperature conditions showed different cell size distribution patterns, which the model was able to reproduce well. Notably, for these days, the model and cell count division rates agreed. This suggests caution in cases where model cell size distributions do not represent observed data well.

The culture division rates exhibited an expected response to growth conditions, typically increasing with temperature (with sufficient light) and increasing with light until saturation. For some temperature and light conditions, though, the division rates of the culture obtained from cell concentrations showed a fairly wide spread of values (e.g., light at 80 μmoles quanta m^{-2} s^{-1}, temperature at 19 °C). This spread can partly be explained from the inclusion of data where cells were adjusting to a change in condition (e.g., just diluted or a recent temperature or light change, see Fig. 2-5). For these days cells would still be acclimating and the division rate of the culture would not be at steady state. The ability of the model to capture much of the observed variation is a further line of evidence for its ability to estimate division rates when cells are adjusting to different environments, as cells continuously do in nature.

2.5.2 Dilution Series Experiments

We used dilution series experiments with natural *Synechococcus* populations as a second line of model evaluation. We used only experiments that demonstrated net

growth rate dynamics consistent with dilution of grazers [54, 29]. The model was applied to cell size distributions measured in undiluted samples (full grazing impact). Model division rates agree well with those obtained from the dilution series technique ($C_b = 0.65$, with a 95% confidence interval of (0.377,0.839)) across a range of division and loss rates (Fig. 2-7A, Table 2.2).

2.5.3 Grazing Assumptions

Our model assumes diel changes in size distributions are only a function of physiological processes. Size selective grazing has the potential to violate this assumption. If *Synechococcus* cells are preferentially ingested according to size, grazing could alter cell size distributions in ways unrelated to cell growth and division. Classes of grazers known to prey on *Synechococcus*, such as nanoflagellates and ciliates, have been observed to be selective across wide size ranges [32, 49], but it is unknown whether this extends to the smaller size differences among cells of a single species growing and dividing.

We found that division rate estimates from the model were not significantly different between the undiluted bottle (higher grazing pressure) and the most-diluted bottle (presumably lower grazing pressure), thus supporting the assumption that size-selective grazing is not important (Fig. 2-7B). We note that the dilution series experiments included in this comparison took place mainly in autumn (only one summer day was included), so we cannot rule out the possibility that grazing effects might occur at other times of year (i.e., grazer community with different prey selection capabilities).

These experiments show that the approach is robust across a range of grazing pressures (Table 2.2). While a few days had low grazing pressure (grazing rate $\leq$ 0.08 d^{-1}), most had relatively higher rates (> 0.25 d^{-1}). These higher grazing rates tended to occur on the second day of incubation, possibly due to grazer reproduction and/or shifts in prey preferences. The agreement in model division rates between bottles in experiments with higher grazing pressure supports our assumption that the model estimates are independent of grazer activity.

2.5.4 Division rates of a natural *Synechococcus* population

The seasonal cycle of *Synechococcus* cell abundance at MVCO is characterized by wintertime low concentrations of a few hundred cells mL^{-1} and summertime levels that can exceed 10^5 cells mL^{-1}. Overlaid on this seasonal pattern are abundance

changes that occur on the timescale of a few weeks; in some cases cell abundance can change by 10-fold (see late summer months in Fig. 2-8A).

Application of the model to the time series of cell size distributions at MVCO in 2008 (Fig. 2-8B) reveals distinct seasonality in division rate. Division rate is very low in winter months and begins to increase almost linearly during spring. Division rate is highest during the summer months and slowly declines in autumn back to wintertime low values. The low rates in winter (0-0.2 d^{-1}) suggest physiological limitation. In summer, however, division rates are the highest observed throughout the year (0.7-1.0 d^{-1}) and suggest that the *Synechococcus* population is not experiencing much resource limitation. The rates produced by the model are consistent with known *Synechococcus* division rates [10, 114, 67], and maximum rates are similar to those observed for the *Synechococcus* strain isolated from this location (Fig. 2-6).

2.6 Discussion

Knowledge of division rates is crucial to understanding drivers of abundance change of phytoplankton populations. A major challenge is not only to partition changes in abundance into growth and loss processes, but also to have this information at a resolution that is relevant for the organism under study. For populations that can be identified and measured at high frequency, the method we have presented makes it possible to obtain daily division rates. We have conducted the first evaluation of the model's ability to estimate division rates from cell size distributions, and find it is able to do so accurately for coastal *Synechococcus*. With this approach, we obtained an unprecedented record of daily division rates over an annual cycle at MVCO. This provides insight not only into the environmental factors that may regulate division rate but also the quantitative role of cell division in producing changes in cell abundance.

Synechococcus division rates show a distinct seasonal cycle (Fig. 2-8). Low values in winter are most likely due to temperature or light limitation. In temperate regions, temperature seems to be a driving force in shaping *Synechococcus* abundance [58, 1, 60]. Waterbury et al. [114] suggested temperature as a limiting factor for *Synechococcus* in this region as initiation of the spring bloom did not occur until water temperature reached 12°C. Our observations support this hypothesis as weekly averaged division rates have a strong positive correlation with temperature during the first half of the year (Fig. 2-9C) when they are only weakly related to light (Fig. 2-10). The increase in division rate begins around mid-April, when water temperature reaches ~8 °C and continues to increase until July, when the highest water temperatures oc-

35

cur (∼22 °C). During the second half of the year, division rates are consistently lower than those encountered in spring for the same temperature. This suggests that other factors limit division rate during this time. Nitrate concentrations are typically low (undetectable or $< 1\mu$M) at MVCO, but higher values are encountered intermittently in fall and winter, likely associated with storms or other mixing processes. High division rates in summer do not indicate extensive physiological limitation, and given undetectable nitrate concentrations, this suggests rapid turnover of nutrients.

With net change in abundance and division rates quantified, we can calculate bulk loss rates as the difference between these quantities (Fig. 2-8C). These loss rates reflect both mortality (e. g., predation, viral lysis) and the net balance of immigration and emigration due to processes such as advection and mixing. Comparison of weekly averaged division and loss rates (Fig. 2-9) illustrates how the combination of division and loss processes produce the observed changes in cell abundance. During the winter and summer months, cell abundance is nearly constant over time scales of several days to weeks, so new cells produced from division must be balanced by losses. This means that loss rates are low in winter and high in summer. Calculated loss rate tends to be correlated with division rate, suggesting that losses are mainly biological in nature rather than associated with advection or mixing of patchy cell distributions. If advection or mixing were dominant, it is unlikely that losses due to these processes would match division rate in magnitude. For *Synechococcus*, the majority of the biological loss term is most likely due to heterotrophic grazers rather than viral lysis [112, 9, 110]. Grazers of picophytoplankton are capable of responding rapidly to increases in prey concentration, as their own division rates can match or exceed that of their prey [52]. A tight balance between division of picophytoplankton and loss by predation has been demonstrated in both open ocean and coastal systems [1, 52, 53, 62, 2, 104]. Our results also demonstrate a tight coupling between division and loss, suggestive of grazing, and reveal that this balance is present over time during the winter and summer months.

At MVCO for 2008, *Synechococcus* population abundance underwent roughly a 1000-fold change during its spring bloom. As required for this bloom to occur, division rate was higher than loss rate during this time (Fig. 2-9A). Compared to the overall magnitude of the division and loss rates themselves, the difference between these two rates, however, was small. The mirror situation occurred in autumn, when cell concentration began to decline. During this time, loss rate was generally higher than division rate, but by only a small amount. An exception occurred during an event in late October, when cell abundance increased by roughly an order of magnitude in the

span of a few days. While we only have division rate estimates for just before and after this period of drastic abundance change, these rates were relatively constant and moderate. This leads us to surmise that this large increase in abundance may have been due to a different water mass with higher *Synechococcus* cell concentration moving into the study area. In contrast to the dramatic multi-day event in October, the large seasonal cell abundance changes appear to result from small (0.1-0.3 d^{-1}), but systematic deviations from 0 weekly averaged net growth rate (Fig. 2-8, solid curve). Spring blooms in temperate locations are usually attributed to temporary escape from predation, especially for the larger eukaryotic phytoplankton. For the *Synechococcus* population at this site, this does not appear to be the case; high biological loss appears to be a constant attribute of the system. The prolonged *Synechococcus* spring bloom is a result of increasing division rates (associated with increasing water temperature) during a 3-month period that allow for a small, but persistent, positive difference over the ongoing losses. This emphasizes the enormous effect that small differences between division and loss processes can have in the evolution of coastal picophytoplankton blooms.

The calculated loss rates presented here also raise interesting questions surrounding the grazing community. Currently, we do not know the identities of *Synechococcus* predators at MVCO. The same type of grazer may persist over the seasonal cycle or different grazers could be selected for by changing *Synechococcus* division rates and abundance. It is important to understand the mechanisms behind the lags in magnitude between loss rate and division rate. Time lags could occur from grazer feeding thresholds or possibly due to differences in temperature responses. Answers to these questions hold important information on the fate of carbon fixed by *Synechococcus*.

Natural *Synechococcus* assemblages are believed to be composed of multiple *ecotypes* [79, 3], each of which may respond differently to its environment. Our model approach provides a composite division rate for the entire *Syenchococcus* assemblage. While our model incorporates two subpopulations, these may very well represent composites of even more finely divided types. Further research is needed to determine if the model can accurately resolve subpopulation division rates of known mixed ecotypes. This would depend on how well physiological differences between ecotypes are manifest in their cell size distributions. Inferences about the contribution of ecotypes to the composite division rate would require quantitative investigations into relative ecotype abundance with molecular approaches and basic knowledge about the physiology of the ecotypes present (growth response to light, nutrients, temperature).

This analysis of *Synechococcus* abundance change demonstrates how our knowl-

edge of population dynamics hinges not only on measurement and observational capability, but also on innovations in modeling and interpretation. The method presented here fills the gap in our current ability to estimate division rate on relevant time scales over extended periods of time. While our model evaluation and experiments focused on *Synechococcus*, this method could be applicable to other phytoplankton (and possibly bacteria) if high-resolution, taxon-specific, diel changes can be characterized. This is more feasible for certain groups of phytoplankton (i.e., *Prochlorococcus* and *Synechococcus*) than others, but current automated imaging technology may make other taxon-specific characterization feasible [75]. Even if taxon-specific changes cannot be realized, this model formulation may also be useful to quantify bulk or average division rate properties. For instance, a version of this approach has been applied by Dall'Olmo et al. [23] to estimate carbon biomass of phytoplankton and division rates from the diel cycle of spectral beam-attenuation coefficients.

For the *Synechococcus* assemblage at MVCO, this method allowed us to estimate daily division rates over an annual cycle and calculate loss rates at this same resolution. We find that throughout the year, growth and loss processes are tightly coupled. Very high cell division rates (up to 1 d^{-1}) can persist for extended periods with little or no change in cell abundance, and the entire seasonal cycle of abundance (3-order magnitude change) results from small (0.1-0.3 d^{-1}) deviations from 0 net growth. We also document strong evidence that division rates are temperature limited during the first half of the year, while other environmental factors are important later. Temperature regulation of seasonal *Synechococcus* blooms may be a characteristic feature in temperate waters (e.g. [58, 1, 60]), suggesting this abundant picophytoplankter may be especially responsive to on-going impacts of climate change. The knowledge gained from this approach contributes to our understanding of *Synechococcus* dynamics and sets the stage to further examine the role that these organisms play in ecological and biogeochemical cycles.
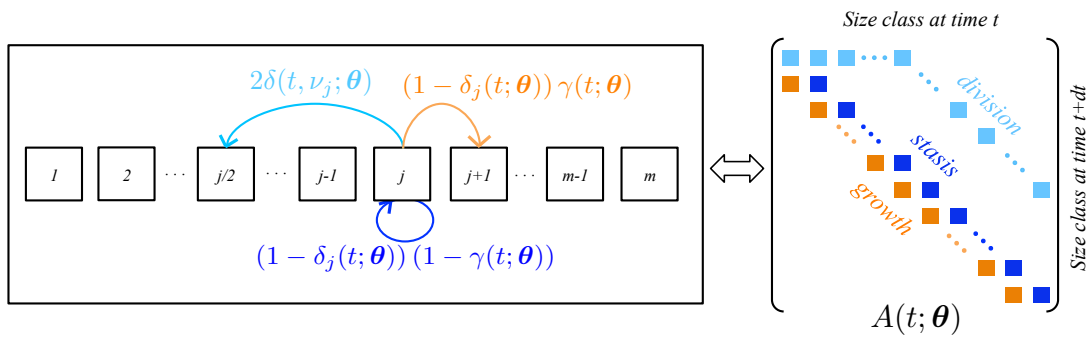
Figure 2-1: Schematic representation of cell size transitions that can occur within one time step, dt, and how the transitions are represented in the matrix, $\mathbf{A}(t; \boldsymbol{\theta})$. The transitions are division (represented in light blue), growth (orange) and stasis (dark blue).
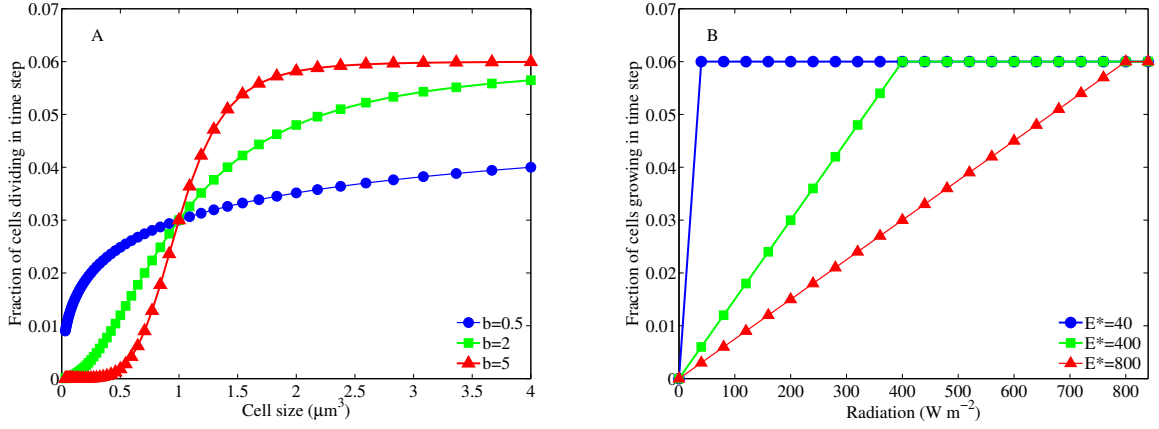
Figure 2-2: Division, δ, and growth, γ, functions with δ_{max} and γ_{max} held constant, but with different values of the respective shape parameters, b and E^*.
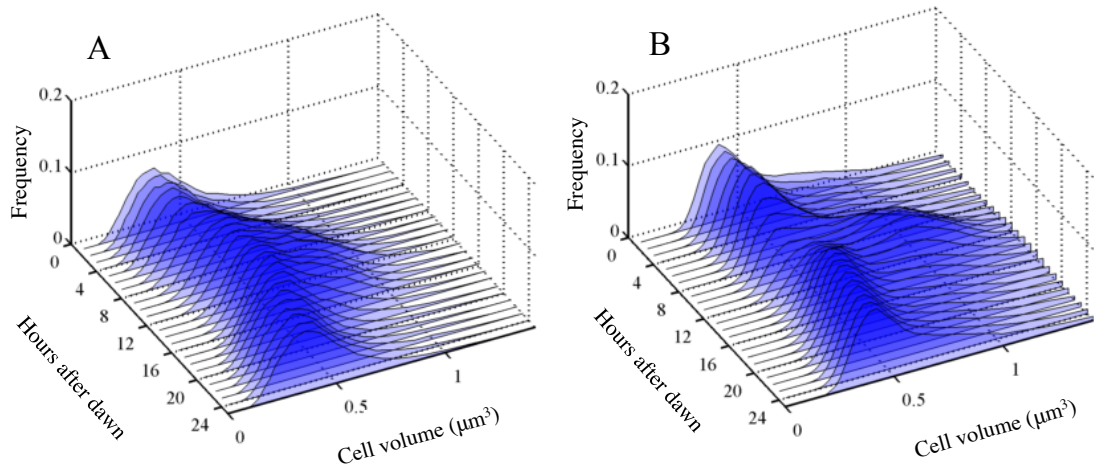
Figure 2-3: A. Observed hourly cell size distributions on June 16th, 2008 at MVCO with estimated division rate of 1.12 d^{-1} from the model. B. Cell size distributions from *Synechococcus* laboratory culture growing at division rate of 0.69 d^{-1} (note bimodal distributions for middle hours of the day).
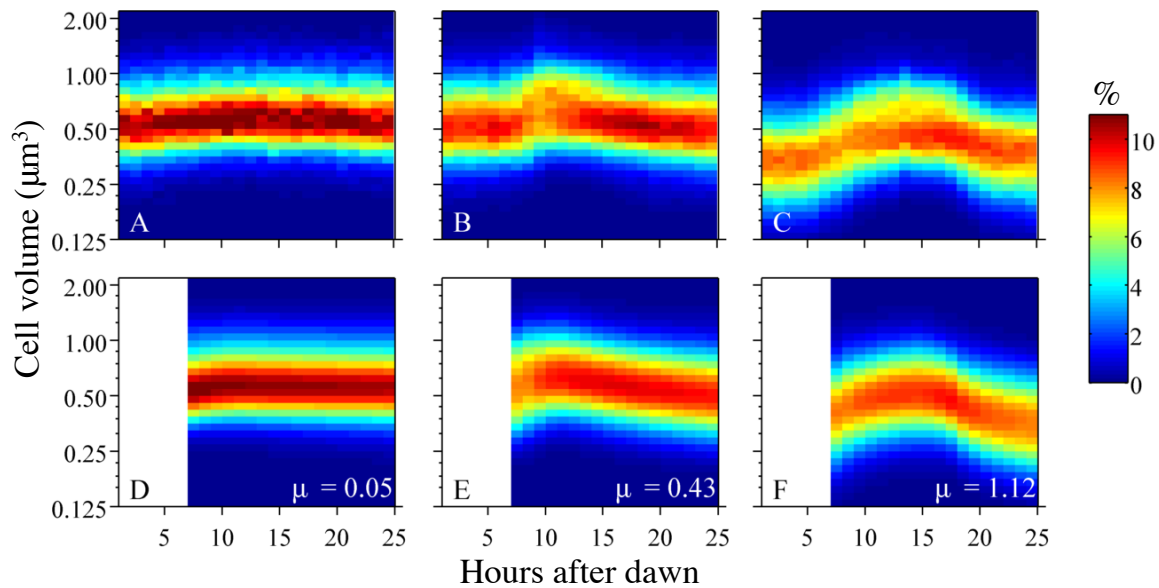
Figure 2-4: Observed hourly cell size distributions of *Synechococcus* obtained from FCB at MVCO on A) 3 January 2008, B) 25 September 2008, and C) 16 June 2008. The bottom row shows the model produced cell size distributions and division rate (μ) from the maximum likelihood estimates of the parameters for each of the observed days directly above. The blank portion from hours 1-6 in D, E, F reflect the model application starting 6 hours after dawn.
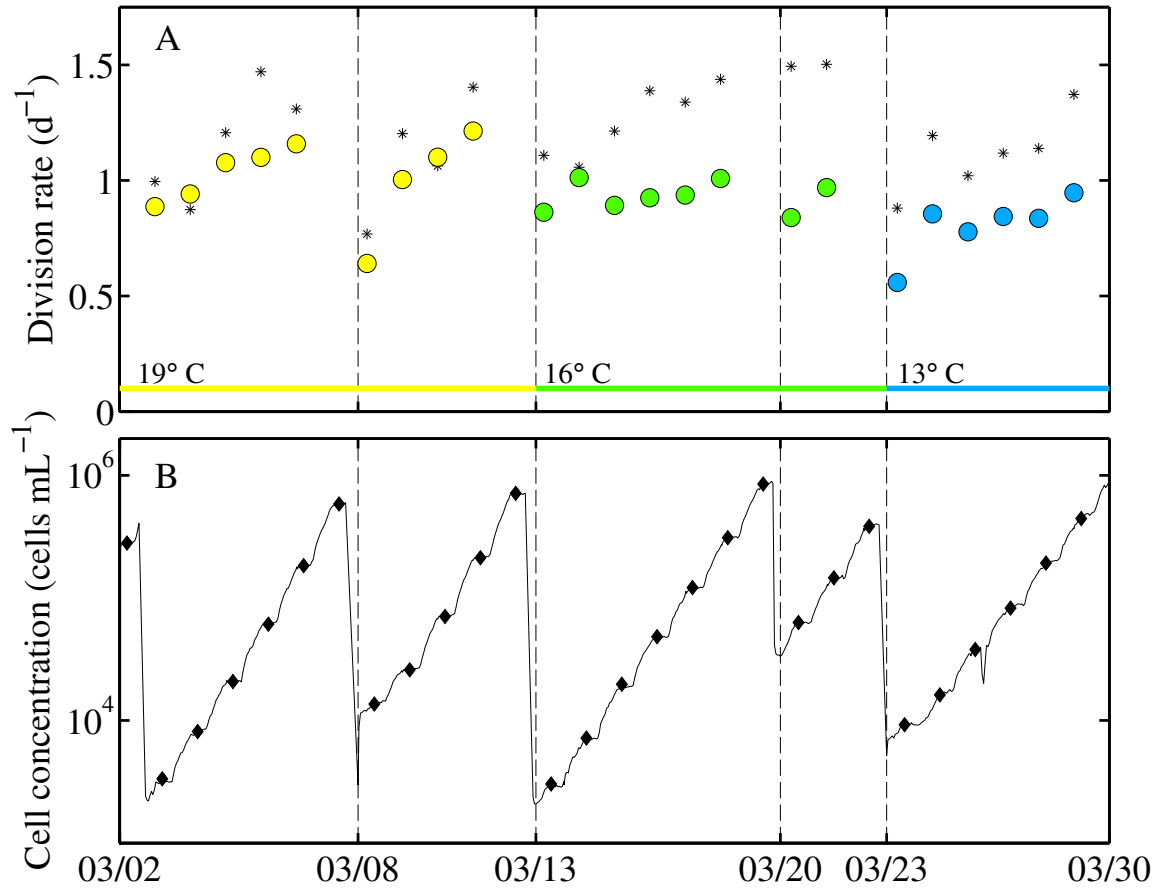
Figure 2-5: A. Daily division rates calculated for successive days of batch growth from the laboratory culture experiments. Colored symbols are division rates calculated from the change in cell concentration as shown in panel B. Color and shape of symbols matches key in Fig. 2-6. Black stars are the model estimates of division rate for each day. B. Cell concentration over the course of batch growth punctuated by dilution with fresh media (indicated by dashed vertical lines). Black diamonds indicate the time and cell concentration used to calculate the division rate in panel A.
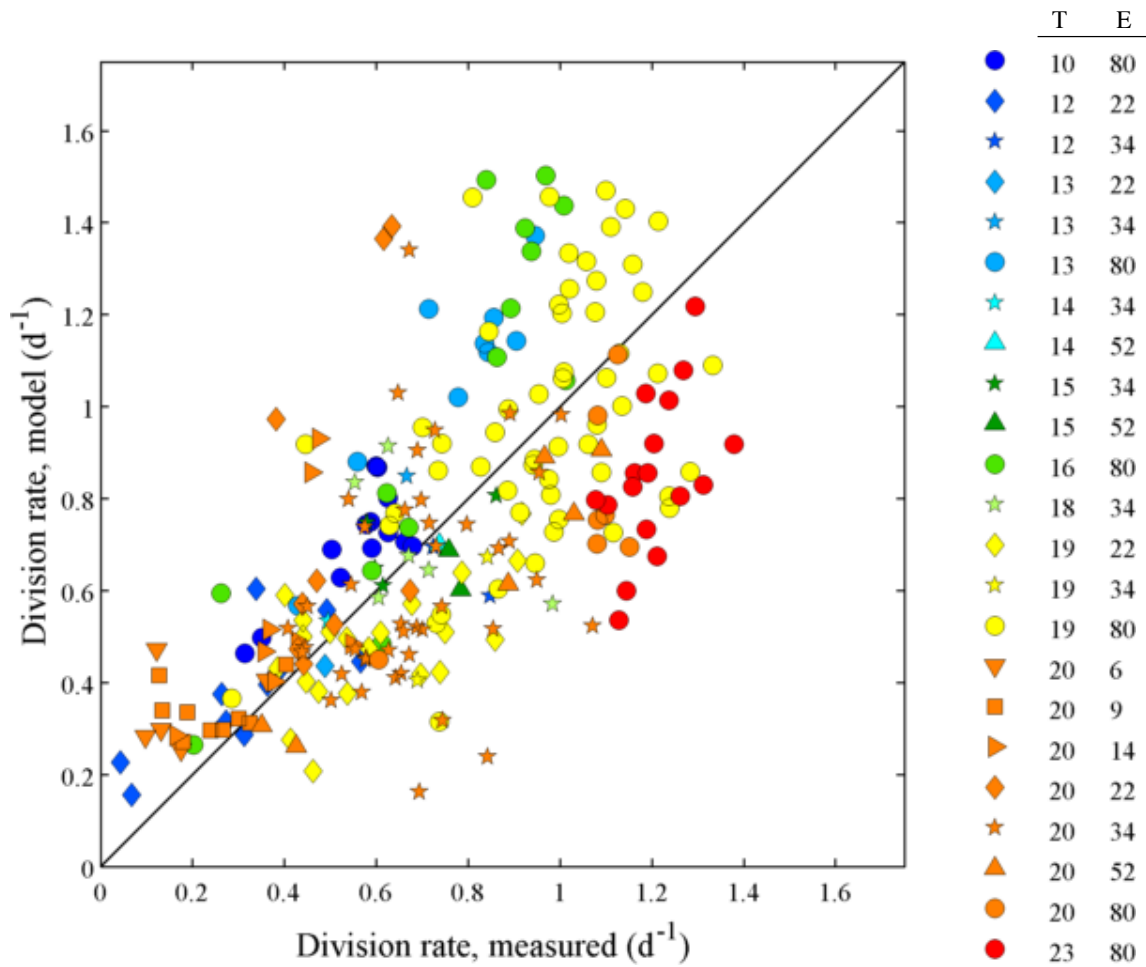
Figure 2-6: Comparison of division rates obtained from the model to division rates determined from change in cell concentration (Eqn. 2.3) for each day in a *Synechococcus* culture grown over a range of light and temperature. Each point is an individual day. T is temperature (°C); E is light (μmol quanta m^{-2} s^{-1}). Color of points indicates temperature and marker symbol indicates light level.
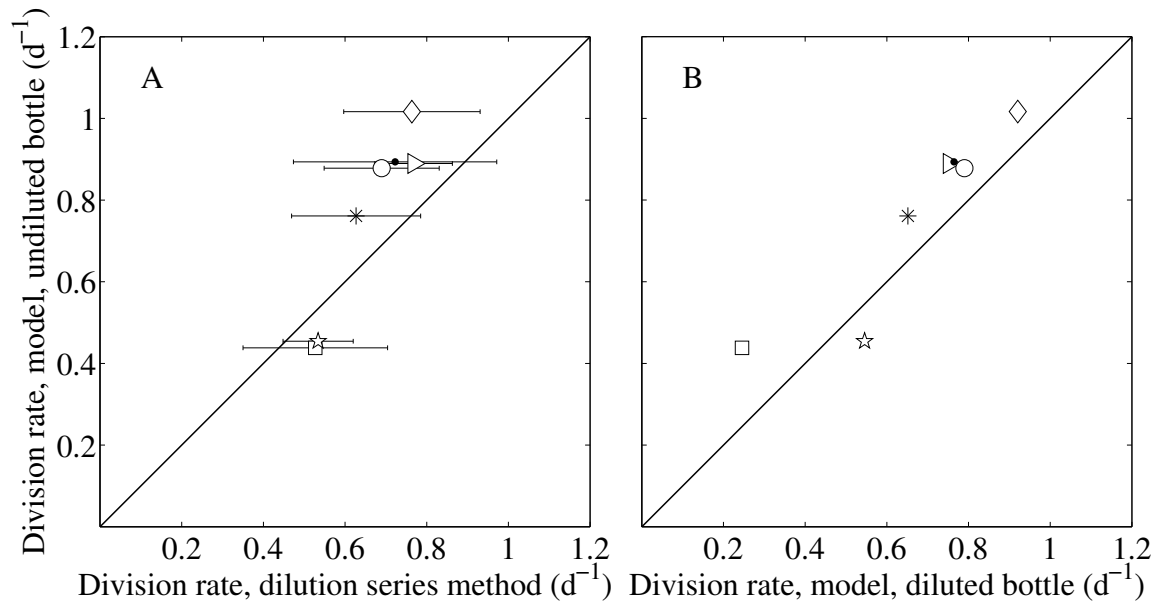
Figure 2-7: A. Relationship between division rate from the model and the dilution series method for assemblages of *Synechococcus* in Woods Hole Harbor water. The model was applied to observations from the undiluted incubation bottle. Each symbol indicates the date of the dilution series experiment (see Table 2.2). 95% confidence intervals are shown for the division rate estimate obtained from the replicated dilution series. B. Relationship between division rate from the model applied to cell size distributions from the undiluted bottle and from the most diluted bottle of each dilution series experiment. The undiluted bottle should have full grazing impact, whereas the diluted bottle should have lower grazing pressure.
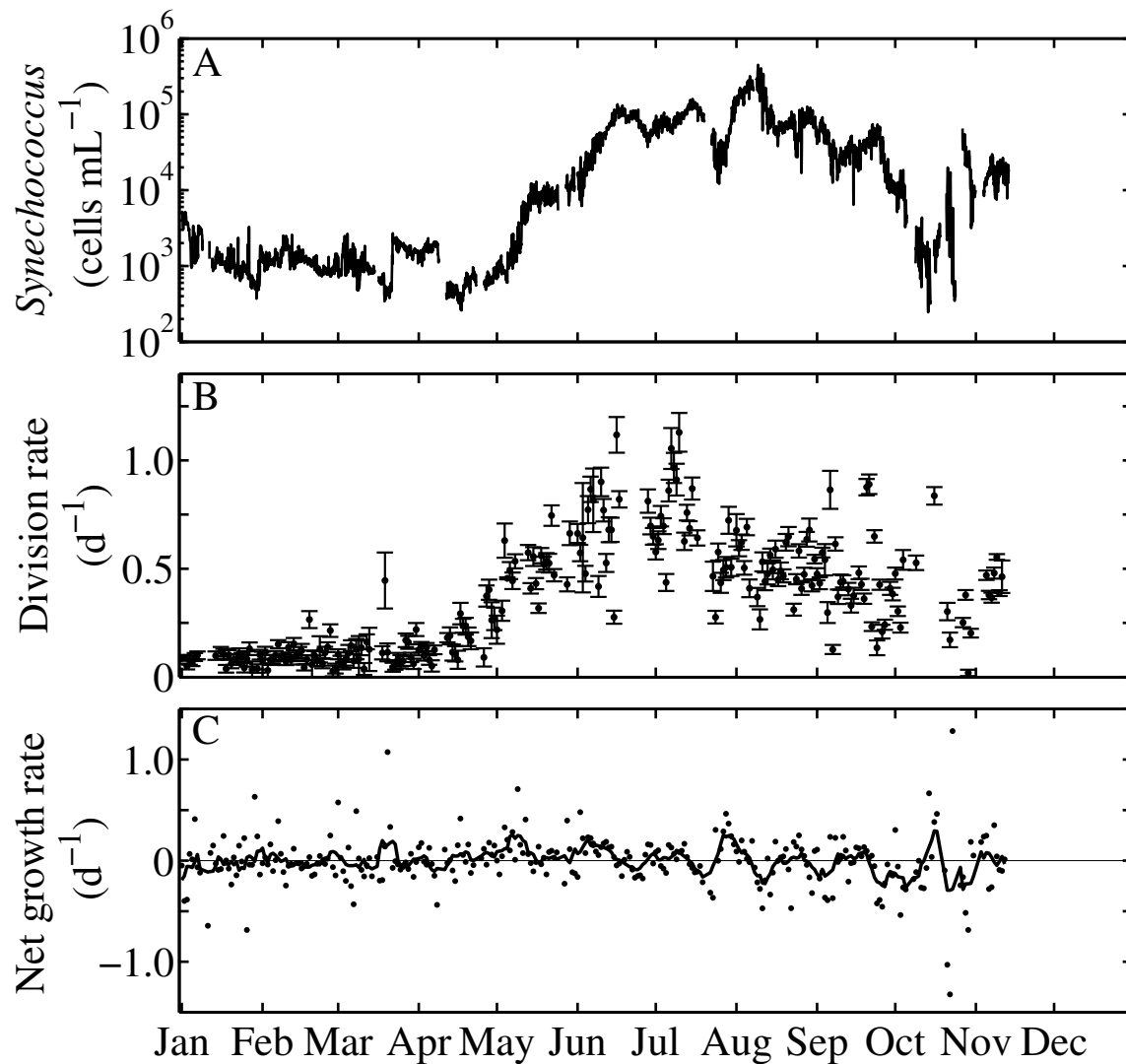
Figure 2-8: A. Hourly *Synechococcus* cell concentration for 2008 at MVCO. B. Daily division rate and 95% confidence intervals from the model. Due to gaps in either FCB data or light data, only 236 days could be considered as input for the model. Confidence intervals reflect the uncertainty in the estimated division rates associated with sampling variability (presuming the model is correct). C. Daily net growth rates obtained with Eqn. 3 and daily averages of smoothed cell concentration (panel A, with a 24-h running mean to reduce tidal effects). Black curve is a 7-day running mean.
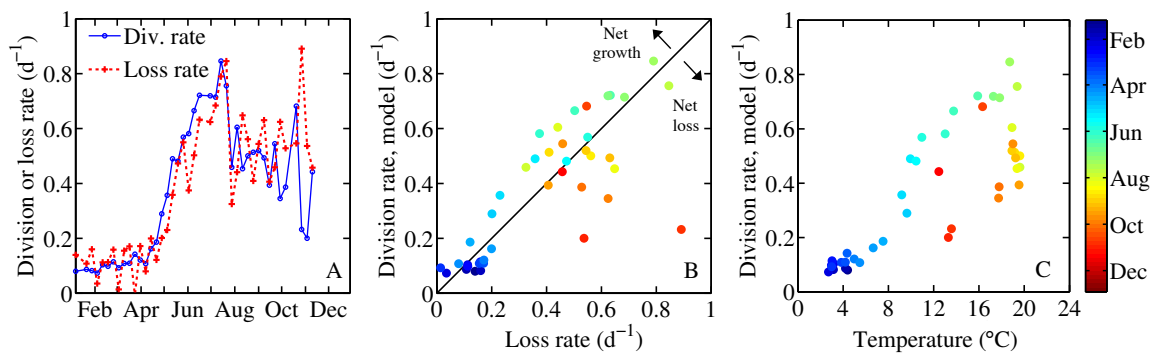
Figure 2-9: A. Weekly averaged division rates, obtained from the model, and computed loss rates for *Synechococcus* at MVCO for 2008. Daily loss rates were calculated by subtracting net growth rate (Fig. 2-8C) from model-produced division rate. B. Comparison of weekly averaged loss and division rates. C. Relationship between weekly averaged division rates and temperature. In B and C symbol color denotes time of year.

Figure 2-10: A. Relationship between weekly averaged division rates and temperature for the first half of 2008 (January 1 - June 30) B. Relationship between weekly averaged division rates and mean incident radiation for the first half of 2008. Color of symbol denotes time of year. Division rates are more strongly correlated with temperature (R = 0.97) for this period than for light (R = 0.68), suggesting that division rate is limited by temperature during this time.

Table 2.1: Variables, constants, and parameters for the matrix model applied to coastal *Synechococcus*.

Item	Definition	Range or Value	Units
$\mathbf{n}(t)$	number of cells in each size class		
$\mathbf{w}(t)$	proportion of cells in each size class		
$N(t)$	total number of cells at each time step		
$\hat{\mathbf{n}}(t)$	number of observed cells in each size class		
$\hat{N}(t)$	total number of observed cells at each time step		
$\mathbf{A}(t;\boldsymbol{\theta})$	projection matrix		
$E(t)$	observed incident radiation		W m^{-2}
t^*	data starting hour and division start time	6	hours after dawn
ν_j	cell size (volume)	$(2^{-5}, 4)$	μm^3
Δ_ν	spacing between size classes	0.125	
ν_{min}	smallest cell size	2^{-5}	μm^3
m	number of size classes	57	
$\boldsymbol{\theta}$	parameter vector		
μ	division rate estimate		d^{-1}
$\delta(t, \nu_j; \boldsymbol{\theta})$	division function		
$\gamma(t; \boldsymbol{\theta})$	growth function		
γ_{max}	maximum growth fraction	$(0, 1)$	
E^*	growth function shape parameter	$(0, \max(E(t)))$	W m^{-2}
δ_{max}	maximum division fraction	$(0, 1)$	
b	division function shape parameter	$(0, 15)$	
ψ	fraction of cells in first subpopulation at t^*	$(0, 0.5)$	
$\bar{\nu}_\ell$	mean cell size of a subpopulation at t^*	$(0.68, 2.38)$	μm^3
σ^2	cell size variance of both subpopulations at t^*	$(0.125, 1.75)$	μm^3
s	precision parameter, Dirichlet-multinomial distribution	$(0, \infty)$	

Table 2.2: Date, symbol, dilution fractions, starting cell concentrations, division rate calculation method, extrapolated division rate, confidence intervals, and grazing rate of each dilution series experiment included in the final dataset. The first and second days of 48-hour incubations are indicated separately. The symbols match those in Fig.2-7

Date	Symbol	Day in bottle	Highest dilution	Starting concentration (cells mL^{-1})	Estimation method	Growth rate (d^{-1})	Confidence interval ($\pm$ d^{-1})	Grazing rate (d^{-1})
June 23	●	2	0.1	278,000	2 phase	0.72	0.25	1.22
Oct 18	$*$	1	0.2	1,674	1 phase	0.63	0.16	0.26
Oct 19	☆	2	0.2	2,409	1 phase	0.53	0.09	0.33
Oct 22	○	1	0.2	893	1 phase	0.69	0.14	0.08
Oct 23	△	2	0.2	1,648	1 phase	0.77	0.09	0.26
Oct 24	□	1	0.2	1,038	1 phase	0.53	0.18	0.05
Oct 25	◇	2	0.2	1,667	2 phase	0.76	0.17	0.28

Chapter 3

Time series investigation of *Synechcococcus* population dynamics

3.1 Introduction

Phytoplankton are important primary producers; they form the base of marine food webs and play critical roles in biogeochemical cycles. Phytoplankton abundances, especially in temperate regions, can undergo large changes over the course of a seasonal cycle. It is important to understand the drivers of these seasonal changes and the extent to which seasonal patterns can vary. However, in order to understand these population dynamics, it necessary to be able to observe individual populations, as taxa will differ in their responses to environmental forcing. It is also necessary to estimate their rate processes at the appropriate time resolution.

Dynamics at the microbial level occur quickly. Division and loss rates are on the order of 1 d^{-1}, acclimation to light and nutrient environments can occur within the span of hours [26], and viral infection and lysis can occur in less than 24 hours [63]. While larger seasonal patterns should have systematic drivers, short term variation will also affect the trajectory of these patterns and it is important to be able to observe this. It is also equally necessary to observe this variation over extended periods of time. This allows investigation into not only the repeatability of patterns or variation, but also characterization of longer term trends.

For the picophytoplankter *Synechococcus*, bottom up control by temperature limitation and top down control by heterotrohpic grazers are thought to be some of the most important factors affecting cell abundance. Strong relationships between division rate and temperature and abundance and temperature have been demonstrated

for *Synechococcus* in temperate regions [114, 1, 58, 60, 109]. Studies investigating grazing on *Synechococcus* over seasonal cycles have demonstrated that heterotrophic grazers are a signification source of mortality and, depending on the season, can control *Synechococcus* abundances [1, 109, 8]. Our investigation of one annual cycle of the *Synechococcus* population at the Martha's Vineyard Coastal Observatory (MVCO) also found that temperature and loss rates are important. With data collected from the automated flow cytometer, FlowCytobot, we were able to obtain data on the *Synechococcus* population with hourly resolution. With time series of cell size distributions and a matrix population model (described in Chapter 2), we estimated daily in situ division rate for the *Synechococcus* population. From these observations, we found that the spring bloom (a 3-order of magnitude increase in cell concentration) results from an increase in division rate associated with increasing temperature. We also found that loss processes were important over the entire annual cycle; division rate was almost always tightly coupled to loss rate. Hence the large seasonal changes in cell abundance were the result of times when net growth deviated only slightly from 0 (0.1-0.3 d^{-1}). The large spring bloom, while triggered by release from temperature limitation, was also regulated by loss processes, influencing the net rate of increase and duration. We now investigate seasonal dynamics of additional annual cycles and determine whether or not the patterns observed for 2008 occur consistently. FlowCytobot has been deployed at the MVCO since 2003, with full year observations beginning in 2007. Application of the matrix population model to the cell size distributions results in a matching time series of daily in situ division rate estimates. These 11-year time series offer an unprecedented look into the dynamics of *Synechococcus*. We find striking repeatable seasonal patterns in both *Synechococcus* cell abundance and cell division rates, as well as calculated loss rates. While these broad seasonal patterns consistently occur, we also observe variation in the timing, magnitude, and other dynamics of these quantities across multiple years. We also observe substantially higher frequency variation that occurs on the scale of weeks to months. We not only explore the relationships among cell abundance, division rate, and environmental factors to understand the causes of the seasonal patterns, but also the interannual variations and multiyear trends. We find evidence that certain features of the *Synechococcus* seasonal cycle are changing systematically as water temperatures warm on the decadal scale.

3.2   Materials and Methods

3.2.1   Study site

The Martha's Vineyard Coastal Observatory is a cabled facility that consists of a shore-based station and meteorological mast, an undersea node located at 12 m depth, and a tower structure in 15 m water rising 10 m above sea level. The node is located 1.5 km south of the island (41° 20.19'N 70° 33.38' W) and the tower is located 3 km offshore (41° 19.5'N 70° 34.0' W). FlowCytobot (FCB) is deployed at 4 m mean water depth at the offshore tower. Deployments from 2003 to 2007 were conducted with only one instrument, but after 2007, with the availability of a second FCB, data was obtained from alternating deployments of each instrument. Core measurements at the MVCO facility include a broad range of meteorological and hydrographic data. Measurements of incident radiation are made at the meteorological mast with an Eppley pyranometer and 20-minute resolution data were downloaded from `ftp://mvcodata.whoi.edu/pub/mvcodata/`. For the majority of 2010, light data was unavailable, and this data gap was filled with radiation measurements obtained from a NOAA National Buoy Data Center buoy (station 44008), located southeast from Nantucket (40°30'9" N 69°14'48" W, $\sim$144 km from MVCO). Daily solar radiation was calculated by integrating incident radiation over 24 hours. Temperature was measured continuously at 4 m at the tower with a MicroCat CTD (Seabird Electronics). Any data gaps in this record were filled with temperature measurements taken at the 12-m node and also downloaded from `ftp://mvcodata.whoi.edu/pub/mvcodata/`. These records have been previously shown to be very similar [81]. Bimonthly-to-monthly sampling trips allowed collection for nutrient analysis. Water samples were collected on board the R/V Tioga using a rosette sampler with water samples taken at discrete depths, usually 2, 6, 10 and 14 m. Samples for nutrient analysis were immediately filtered through a 0.2 μm Sterivex® filter into acid-washed vials and frozen at -20 °C. Samples were analyzed for phosphate, ammonium, silicate and nitrate + nitrite by standard autoanalyzer techniques at the Woods Hole Oceanographic Institution Nutrient Analytical Facility (Woods Hole, MA).

3.2.2   FlowCytobot

Details of the design and performance of the automated, submersible, flow cytometer, FCB, are described elsewhere [74]. Briefly, the instrument uses a 532-nm solid state laser for excitation and is able to detect individual cell forward and side light scattering

and fluorescence at 575 and 680 nm. FCB includes pairs of linear amplifiers set to different gains to extend the dynamic range, signal processing electronics, and a computer for sample control and acquisition. Data analysis and enumeration of *Synechococcus* cells were as described in [98]. *Synechococcus* are discriminated from other phytoplankton and particles by their characteristic orange fluorescence from phycoerythrin and small light scattering signals. FCB measures side angle scattering, which has been calibrated to estimate cell volume [74]. Polystyrene microspheres (beads) (Polysciences Inc.) of diameter 0.5 μm (polychromatic) and 1.0 μm (red-fluorescing) were measured as reference particles every $\sim$ 20 h during deployments.

3.2.3 Division rate estimation

To estimate daily population division rate, we applied the two-subpopulation matrix population model (described in Chapter 2) to hourly binned cell size distributions obtained from FCB and radiation data. Constants and parameter constraints were the same as listed in Table 2.1. To find the maximum likelihood estimates of the parameters, we utilized a constrained nonlinear optimization routine offered in MATLAB (fmincon). Due to the complexity of the parameter space, we initialized the solver with at least 40 random starting points. Best fit parameter values were chosen if they satisfied the following criteria: 1) At least 5 solver runs resulted in parameter estimates for which maximum likelihood values were within 0.2 distance from each other, and 2) parameter values were within 5% difference from each other or within a specified absolute error tolerance. If the initial starting points did not result in parameter values that satisfied these criteria, another 40 random start points were chosen up to 200 solver runs. The best fit parameters were used to estimate a division rate as described in Chapter 2.

3.2.4 Data analysis

Cell concentration was smoothed with a 48-hour running mean to reduce tidal effects. We calculated net growth rate from this data (μ_{net}), as follows:

$$\mu_{net} = \ln\left(\frac{\bar{N}(t_d + 24)}{\bar{N}(t_d)}\right),$$

where t_d is the hour of dawn of each day, $N(t)$ is the smoothed cell concentration at hour t, and $\bar{N}(t_d) = \frac{1}{3}\sum_{j=-1}^{j=1} N(t_d + j)$. This gives an average cell concentration in the 3-hour interval surrounding dawn. Net growth rates calculated using these

average dawn-to-dawn values match time period used for division rate estimation. Loss rates were calculated by subtracting net growth rate from division rate for each day available in the dataset.

Daily-resolved annual climatologies were calculated by averaging values across available years. Weekly climatologies were constructed by first averaging values belonging to each week in the year and then averaging over the same week for all years. Anomalies at daily resolution were constructed by subtracting the climatological value from the data for the corresponding year day. Correlations between different anomalies and significance tests were carried out in MATLAB, and assume a Student's t-distribution to calculate the test statistic.

To determine the point at which division rate began increasing during the spring, we fit a second order polynomial to the data between January 1 and June 1 of each year. We chose the point at which the polynomial crossed the threshold of 0.2 d^{-1} as the date at which division rate began increasing (Fig. A-16). Similarly, to identify a point at which the increase in division rate began to slow and reach maximum values, we fit a second order polynomial to division rate estimates between May 1 and July 1 of each year. The point at which the derivative of the polynomial crossed a threshold of 0.003 was chosen as the point at which division rate stopped increasing for spring (Fig. A-17). We estimated the end to the spring bloom at the point at which net growth rate dropped below -0.05 d^{-1} for two consecutive weeks.

3.3   Results

3.3.1   Environmental Conditions

As expected for a temperate location, solar radiation and temperature vary over a wide range throughout the annual cycle (Fig. 3-1, 3-2). Daily radiation ranges from a minimum of $\sim$ 5 MJ m^{-2} in late December to $\sim$ 25 MJ m^{-2} in late June. The timing of minimum and maximum temperature is, as expected, offset from the pattern of radiation. Average minimum temperature is $\sim$ 2°C during mid-February and reaches a maximum of $\sim$ 20°C in mid-August. There is a large amount of variance in the radiation climatology (Fig. 3-2B), while the temperature climatology demonstrates less variance (Fig. 3-2C). Temperature values in winter, however, demonstrate higher variance than values for the rest of the year.

Measured nutrients (nitrate + nitrite, ammonium, inorganic phosphate and silicate) typically display much less seasonality (Fig. 3-3). Nitrate + nitrite concentra-

tions are typically low and ≤ 1 μM, with higher values sometimes found in fall and early winter. Ammonium is usually ≤ 3 μM and inorganic phosphate is typically $\leq$ 10 μM. Given that the resolution of nutrient data is lower than temperature or radiation data (bimonthly-monhtly vs. 20 min., respectively), exploration of the effect of nutrient availability on population dynamics are limited.

3.3.2 Cell Abundance

Synechococcus cell abundance is characterized by distinct seasonal features (Fig. 3-4, 3-5). Cell abundance in late winter through mid-spring (February-April) is typically low, hovering between 100 and 1000 cells mL^{-1}. A systematic increase in cell abundance from April through mid-June characterizes the spring bloom. During this time, cell concentration undergoes a two-three order of magnitude increase up $\sim 10^5$ cells mL^{-1} in the span of roughly 2 months and reaches its highest concentration for the entire year. Cell concentration remains relatively high (above 10^4 cells mL^{-1} through mid-summer into early winter (July-December), but throughout this time is slowly declining. During winter, the rate of decline is much sharper until cell concentrations return to late winter and early spring low values.

The fast increase and slow decline of cell abundance in the spring and summer, respectively, are illustrated by the seasonal pattern in net growth rate (Fig. 3-6C, D). From the climatology, the spring bloom is the only time of year when net growth rate is consistently positive for longer than a few days (0.15 d^{-1} on average). During summer, net growth rate hovers just around zero (or slightly below) during summer and then typically remains negative for fall and winter. Variance in the net growth rate climatology is usually much larger during the summer through fall (Fig. 3-6B,E). At this time, cell concentration can undergo large changes (up to an order of magnitude) on short time scales (a few weeks). These shorter duration increases and decreases are observed throughout the seasonal cycle, but are most dramatic for this season.

3.3.3 Cell Volume and Fluorescence

Synechococcus cell volume (calculated from bead-normalized side scattering measurements of individual cells) follows a distinct seasonal trend as well (Fig. 3-7). Cells have a larger average cell volume in late fall through winter and into early spring (0.5 - 0.75 μm^3) than in summer (0.25 - 0.5 μm^3). Bead normalized cellular fluorescence also follows a seasonal trend. Fluorescence decreases between January and June, at which point it begins to increase until October. For the rest of the year, it remains

relatively constant.

3.3.4   Division Rate

Division rate also follows a seasonal pattern (Fig. 3-6A,D, 3-8A). Values are consistently low (~ 0.1 d^{-1}) in January and February. In March, division rate just begins to increase and by April increases almost linearly until June, at which time climatological values reach a maximum of 0.9 d^{-1} (although higher values are observed in individual years). Division rates remain relatively high in the summer, but slowly decline during the time from July through October. At this point in the year, division rate undergoes a sharper decline to winter time low values. Variance of division rate, is much less during winter and during the spring bloom than in summer or fall. In general, division rates remain moderate to high in summer and fall, but we observe brief stretches (few days) during these seasons when they are much lower (0.2-0.5 d^{-1}).

3.3.5   Loss Rate.

Calculated loss rates follow a seasonal pattern very close to that of division rate (as required since net growth rate hovers near zero). Losses closely track the division rate in magnitude over time: increasing and decreasing when division rate does the same (Fig. 3-6E,F, 3-8A,B). The variance in loss rate is large throughout most of the year, even in winter when cell concentration and division rates show less variance.

3.4   Discussion

3.4.1   A general framework

We now explore the relationships between cell abundance, division rate, and prevailing environmental conditions to understand the causes of the seasonal pattern of cell abundance and also explore interannual variability. From the climatologies, we can construct a framework for how these factors relate to one another and then use this as a starting point from which to investigate variations in the seasonal cycles and differences between years.

Relationships between division rate and environmental factors

Division rate is a critical metric of the physiological state of the population at a given time. Low division rates indicate physiological limitation and the relationships between division rate and environmental factors allow us to make inferences about which factors restrict division rates. Over the average annual cycle, division rate is related to both temperature and light. For the first half of the year, from January through the start of June, division rate is correlated with temperature (Fig. 3-9B, Fig. 3-10, Fig. A-14A), while there does not appear to be a systematic relationship with radiation (Fig. 3-9C, Fig. A-14B). This supports the conclusion that division rate during this time of year is temperature limited. Light levels are still low during winter, but increasing solar radiation does not correlate with an increase in division rate for this period. A possible exception is when temperature is above 5 degrees and solar radiation is still below 18 MJ m^{-2}; light may have an influence during this period, but both light and temperature are changing rapidly, so their effects are difficult to separate. Division rate increases linearly once the water temperature has reached 5-7 °C (Fig. 3-9B). For the summer and early fall (June through September) division rate is usually greater than 0.6 d^{-1}, and does not demonstrate any pattern with light or with temperature (Fig. A-14B,E). These rates, while still moderate to high in value, are lower than at the very start of the summer (June), possibly indicating other physiological limitation (i.e., nutrient availability). From October and through December, division rates systematically decrease and show a strong correlation with both temperature and light (Fig. 3-9B,C, Fig. 3-10, Fig. 3-11). Division rates were much higher in the spring for the same temperatures, suggesting that light limitation may be the controlling factor in the fall. These relationships can be readily observed as a three dimensional plot between division rate, temperature, and light (Fig. 3-12).

Relationships between division rate and loss rate

We now have hypotheses about the main controls on how fast cells are added into the system at different times of year: temperature limitation during late winter and spring, possible nutrient limitation during summer, and light limitation during late fall and early winter. The general pattern of cell abundance, however, is determined by the balance of the rates at which cells are added to the system and the rates at which cells are removed. Calculated bulk loss rates allow us to investigate factors that may control the rate at which cells are removed. We find that, in general, loss rate closely tracks division rate in magnitude over the course of the year (Fig. 3-13, 3-14).

The main source of loss for the *Synechococcus* population are likely to be heterotrophic grazers. This broad term encompasses the heterotrophic nanoflagellates, ciliates, and dinoflagellates, all which have been reported to feed on *Synechococcus* [21, 15, 85, 20, 44] and may be important classes of grazers at MVCO. These organisms can quickly react to and take advantage of an increase in prey cells as their own cell division rates can match or exceed those of *Synechococcus* [52]. Losses due to viruses may also play a role, but in general viral lysis is thought to contribute less to *Synechococcus* losses than grazing [112, 8, 110].

While there is generally a very close coupling of loss to division rate, as necessary for cell concentration to increase and decrease systematically, there are times of year when these two rates are not equal (Fig. A-15). The largest imbalance occurs during the spring bloom. Interestingly, this imbalance does not start when division rate begins to increase. From the climatological view of spring, division rate only exceeds loss rate after the water temperature exceeds 8 °C (Fig. 3-13). Initially, division rate only escapes loss rate by a small amount, but then division rate increases more rapidly than loss rate for 2 months. The spring bloom is therefore determined by both the increase in division rate and the trajectory of increasing loss rate. To the extent the losses during spring reflect activity of heterotrophic grazers, the dramatic increase in losses may reflect a combination of grazer activity increasing with temperature and higher ingestion rates as *Synechococcus* abundance continues to rise. Loss rate shows a correlation with temperature once temperature is above 6 °C (Fig. 3-15A), but does not show a relationship with *Synechococcus* cell concentration until May (Fig. 3-15B), when cell concentration reaches around 10^4 cells mL^{-1}.

The slow decline of cell concentration in late summer and fall, as well as the steeper decline in winter, appears to be the result of decreasing division rate coupled to loss rates that just exceed division rate. If losses primarily result from grazers, then loss rate will mostly be determined by prey availability, number of grazers present, and temperature. While division rate is decreasing, cell concentrations are still within the range for which ingestion rate can be relatively high [20], such that the *Synechococcus* population could still be experiencing heavy grazing losses. From September through December, we also find a relationship between loss rate and temperature (Fig. 3-15). While this relationship can be explained by the relationships between division rate and temperature and loss rate and division rate, temperature will likely exert a direct effect on loss rate if the majority of the losses are biological (i.e., heterotrophic grazers). We also find that during this time, loss rate also correlates with *Synehcococcus* cell concentration. This climatological relationship is different for the fall than in

the spring. Loss rates in the spring are higher than losses in the fall for the same *Synechococcus* cell concentrations. This could be a reflection of different loss agents present at different times of year (i.e., grazers vs. viral infection, grazers with different ingestion capabilities or selections).

3.4.2 Variation in seasonal patterns

From the climatologies, we now have a starting framework from which explore and examine the variation among years. We take a closer look at the different seasonal abundance patterns and variations: the spring bloom, summer and fall cell abundance, and the winter decline.

Spring bloom

The spring bloom is identified by a rapid and large increase in cell concentration. However, there can be substantial variation in how the bloom evolves for each year. With the resolution of our observations, we are able to observe that the spring bloom is not simply a straight forward increase in cell concentration. Typically, there are shorter periods of decreases or no change in cell concentration that are overlaid on this mulit-month increase in cell abundance. These patterns can be explained by how both division rate and loss rate vary during the spring (Fig. 3-17).

The conditions at which the *Synechococcus* population division rate begins to increase appears to be relatively consistent throughout the time series. Division rate begins increasing when water temperature is in the range of 5-7 °C (Fig. 3-16A). Division rate anomalies are positively correlated to water temperature anomalies when the temperature is below 10 °C (Fig. 3-18A). In general, division rate continues to increase with increasing temperature, but it is not unusual for division rate to slow, stop increasing, or decline at some point during the bloom (Fig. 3-17). Depending on the temperature, these lower division rates may be the result of temporary light limitation. Anomalies in solar radiation showed a low, but significant, correlation to division rate for solar radiation levels below 8 MJ m^{-2} (Fig. 3-19A). This idea was put forth by Waterbury et al. [114], who observed disruptions in the increase of *Synechococcus* cells during the spring bloom in Woods Hole Harbor following a major storm. These authors also observed changes to net growth rate following rainfall and a lower percentage of dividing cells during consecutive cloudy days. During the period when radiation is still increasing for spring, storms or extended cloud cover could result in light limitation. We sometimes find evidence for this (see Fig. A-23C), but

we also observe lower division rates when light should not be limiting. This suggests other factors, such as nutrient availability may also be important. The shorter time scale variation of division rate could also reflect environmental patchiness, such that favorable growth conditions may not be uniform along the shelf.

The evolution of the spring bloom is complicated by variable loss rates during this time. Loss rates follow a general increase, but also demonstrate declines, stops, or slowed rates of increase. Typically, we find that loss rate is first able to match division rate when the latter just begins to increase (Fig. 3-17). If losses are due to grazers, this suggests rapid response of the grazing community to the change in activity and increase of *Synechococcus* cells. After these initial increases, we find loss rate to be much more variable. These variations could be due to different temperature responses than *Synechococcus*, feeding thresholds, availability of other prey, increase in the predators of the grazers themselves or other factors. Only when these two rates deviate for an extended time (such that division rate becomes higher than loss rate), can cell concentration increase. These periods of sustained positive net growth rate vary from year to year in timing and duration, and could indicate complicated dynamics between *Synechococcus* and their grazers. Some of the variability in losses could also be explained by advection and patchiness. Water parcels containing higher or lower *Synechococcus* cell concentrations or grazer concentrations may have different loss dynamics, and advection of these water parcels past MVCO may result in some of the variability observed on a short time scale.

The end of the spring bloom can be difficult to pin point, as shorter scale variations in cell abundance obscure the general trend of slowed increase in cell concentration. In general, we find that cell division rate reaches a maximum value when water temperature is $\sim 15°$C. The metric we have used as a rough estimate for when the bloom ends (see Materials and Methods) indicates that cell concentration stops increasing on average when water temperature is 15-17 °C and cell concentration has reached $1 - 3 \cdot 10^5$ cells mL^{-1}. This is typically when cell concentration reaches its annual maximum, but occasionally the highest values can occur later in the summer (i.e., years 2008, 2010). It is interesting that *Synechococus* cell concentrations are not able to increase above $\sim 3 \cdot 10^5$ cells mL^{-1}. This may indicate a grazer saturation concentration for which grazers (if only ingesting *Synechococcus*) are no longer food-limited and can achieve their maximum division rates. For nanoflagellates observed in other studies, these end-bloom concentrations appear to be just slightly below ingestion saturation concentrations [20, 109].

Summer and Fall Dynamics

Throughout June to mid-December, cell abundance remains relatively high ($> 10^4$ cells mL^{-1}), but there is a steady decline through and into late fall. From the climatological patterns, we observe that the general decline is a result of decreasing division rates during this period coupled with loss rates that are just higher than division rates. As discussed in section 3.4.1, it is unclear what factor may be limiting division rate in June-September, but the decreasing division rates from October through mid-December are likely driven principally by the seasonal decrease in solar radiation. If we examine the correlations between division rate anomalies and radiation anomalies, we find that division rate shows a low, but significant, correlation to anomalies at levels below 16 MJ m^{-2} (Fig. 3-19). We also fine that division rate anomalies during this time were less highly correlated with temperature anomalies (although correlation was still significant) (Figs. 3-18, A-19) than in the spring. Furthermore, when we examine division rate only as a function of solar radiation (temperature held almost constant), we observe what appears to be light response curves (i.e., saturating response of division rate to increasing radiation levels, Fig. 3-20). These curves illustrate temperature envelopes for division rate as a function of radiation. Notably, when temperature is below 16°C, we find that the majority of division rates fall along the 'light limited' portion of these response curves, further suggesting that light is the main limiting factor during fall.

While solar radiation incident on the ocean surface is changing rapidly in the fall, the light environment that a *Synechococcus* cell experiences will also be affected by the attenuation coefficient and by the rate of mixing within the water column. These variables are likely explanations for why the relationship between solar radiation and division rate for different temperature intervals demonstrate such high variance. Despite the unexplained variance, these curves suggest that division rates are not light limited above ∼10 MJ m^{-2}. Exceptions occur during the times of year when temperature is above 16 °C. Many points appear to fall below a maximum division rate envelope and could indicate nutrient or other physiological limitation during this time. Nonetheless, the variation in temperature and light during fall can explain much of the variation observed in division rate, and further supports our hypothesis that maximum division rate is controlled by temperature, but light limited during this time.

As with the spring bloom, we also observe short timescale fluctuations in cell concentration (with ∼2-3 week period) that can be up to order of magnitude changes. Variations in either division rate or loss rate could result in these dynamics. Specif-

ically, we observe multi-day declines in division rate, which are typically followed by a return to moderately high rates. While this could indicate temporary nutrient limitation, the spareness of our nutrient observations makes this difficult to evaluate. Variations in loss rate also occur, and could result from grazing saturation thresholds, decreases or increases in abundances of grazers and viruses or patchiness and advection. The fluctuations tend to follow fairly consistent patterns of increases and decreases, with peaks in abundance separated on average by two weeks. Predator-prey dynamics have been shown to produce similar oscillations to those observed here [51] and are a likely cause. Viral infection, though, could also result in similar patterns. Natural *Synechococcus* assemblages are diverse, with many viruses thought to be specific to certain *Synechococcus* types [71]. These periodic decreases could be a result of viral lysis of one type of *Synechococcus* or an attack on sensitive types [112] that have temporarily been allowed to increase.

It is unknown how infection by viruses affects the cell size distributions of *Synechococcus* and how such size dynamics would influence the division rate estimates obtained with our matrix population model. During a lytic infection, cells should not be dividing (although division can occur if cells are stressed (i.e. pseudolysogeny) [63]). Cell size, however, could still vary, perhaps during phage assembly or just before bursting. If there is no cell size change during an infection and subsequent burst, then division rate for the cells undergoing infection should be low. However, if the proportion of the population infected (and not dividing) is low, then this may not be reflected in the overall population division rate, as this is a composite estimate. Therefore, it may be difficult to detect viral infection via division rate if only a small percentage is undergoing a lytic cycle.

Winter Decline

The low winter concentrations (100-1000 cells mL^{-1}) are the result of very low division rates (0.05-0.15 d^{-1}) (from temperature limitation) coupled to loss rates that are just higher than division rate (0.15-0.2 d^{-1}) . For the winters we observed, we find two general patterns: either the cell concentration is able to stay relatively high, at $\sim$ 1000 cells mL^{-1} or it continues to drop to a few hundred cells mL^{-1}. Anomalies in cell abundance during winter are correlated with temperature anomalies (as are division rate anomalies) (Fig. 3-18). This suggests that the minimum cell concentration reached is primarily determined by how cold the winter is. During the winter, loss rates are also low, but still highly variable. Grazers could still be active, as psychrophilic nanoflagellates have been observed [19], but the *Synechococcus* population

numbers are also low, suggesting that grazers targeting *Synechococcus* might be food limited. Lytic viruses would also have difficulty sustaining infection under these low *Synechoccocus* concentrations [63]. This potentially points to advection and other physical process as a source of loss or dilution of the *Synechococcus* population with water masses containing lower *Synechococcus* concentrations.

3.4.3 Temperature implications

Mean ocean water temperature temperatures are already rising in the coastal waters around MVCO [72], and are expected to continue warming over the next few decades. Given that *Synechococcus* dynamics are closely linked with temperature, they will likely be impacted by this increase. Certain features of the seasonal cycles may be more affected than others. Because temperature is the underlying trigger of the spring bloom, this suggests that increasing water temperature will have an impact on bloom dynamics. In fact, during the decade we have observed, there is already evidence that *Synechococcus* division rate is beginning to increase earlier in the year. This observation can be linked to the fact that water temperature is warming to 6 °C earlier in the year (Fig. 3-16B). From cell concentration observations, we also find evidence that the the bloom has shifted ~ 20 days earlier over the last decade (Fig. 3-16C). While the exact start date of the bloom is difficult to pinpoint for the reasons discussed above, we can clearly see this shifting by examining threshold cell concentrations that indicate a 'midpoint' in the bloom. We find the time at which cell concentration reaches a certain value, (i.e., 10^4, $5 \cdot 10^4$, or 10^5 cells mL^{-1}) has, on average, shifted earlier. While the exact bloom trajectory is still highly variable, these trends suggest that the underlying mechanisms are sensitive to changes in temperature. The bloom is starting earlier each year and reaching threshold cell concentrations earlier as well. Higher temperatures during winter will also likely result in a higher standing stock of *Synechococcus* at the beginning of blooms. However, increases in temperature may not necessarily translate into higher concentrations for the rest of the year. Division rates in summer are typically lower than those in spring, suggesting that increased temperature at this time will not translate into higher division rates. Given that loss rates at this time are particularly high, any increase in *Synechococcus* production may translate into more turnover into consumers. This has important implications for ecological and biogeochemical cycles that involve *Synechococcus*. We may anticipate that increasing temperatures will result in shortened winter and spring dynamics, and a microbial loop that spins faster during the summer and fall.

3.4.4 Conclusions and Future Directions

Our investigation of *Synechococcus* dynamics has illustrated some of the underlying causes of seasonal population dynamics. We find that temperature is an important control on division rate in the first half of the year. The spring bloom results from release of temperature limitation, but is also determined by the rate of at which losses increase. During the summer, division rates are typically high, indicating little resource limitation. To the extent there is variability in division rates in summer, lack of relationships with temperature or light suggest that other factors (i.e., intermittent nutrient availability) must be at play. During fall, we find evidence that light limitation is the primary driver. The steady decline of *Synechococcus* during this time is a result of declining division rate coupled to loss rates that are just higher than division rates. In winter, both cell concentration and division rate vary with temperature, with higher values found for warmer winter temperatures.

From the high resolution data we also see significant variation in these patterns. Variations in cell concentration appear to result from a complex interplay between *Synechococcus* cell division and losses due to grazers or viral lysis. While this study mainly focused on the larger seasonal scale patterns and interannual variability, high resolution observations allow exploration at very short time scales. Investigations into these shorter time scale processes will be important to understand how short term variation can affect the larger seasonal patterns. The conclusions of this study highlight the need for greater understanding of biological losses. Ideally this would be supported by relevant observations at the same time resolution for which we are able observe *Synechococcus*. We do not yet know the identity of the *Synehcococcus* grazers at MVCO, but once identified, current automated imaging technology (Imaging FlowCytobot [75]) could allow enumeration of target taxa species. Information on predator dynamics is important not only to better understand the regulation of *Synechococcus*, but also to determine the fate of the carbon fixed by *Synechococcus*. Whether or not this carbon is transferred to higher trophic levels depends on the structure of the food web and the role *Synechococcus* grazers play [52].

Variations in cell abundance and division rate may also be influenced by the underlying diversity of the *Synechococcus* assemblage itself. *Synechococcus* is diverse genetically and physiologically, with strains demonstrating differences in nutrient utilization, light acclimation, and temperature responses [68, 80, 83], as well as different palpability to grazers [6, 121]. Differences in environmental and ecological responses between *Syenchococcus* types may be responsible not only for the variation in the season cycles, but also for the general patterns themselves. In the next two chapters,

I explore the diversity of the *Synechococcus* population at MVCO (Chapter 4) and how this composition may affect the abundance dynamics explored here (Chapter 5).
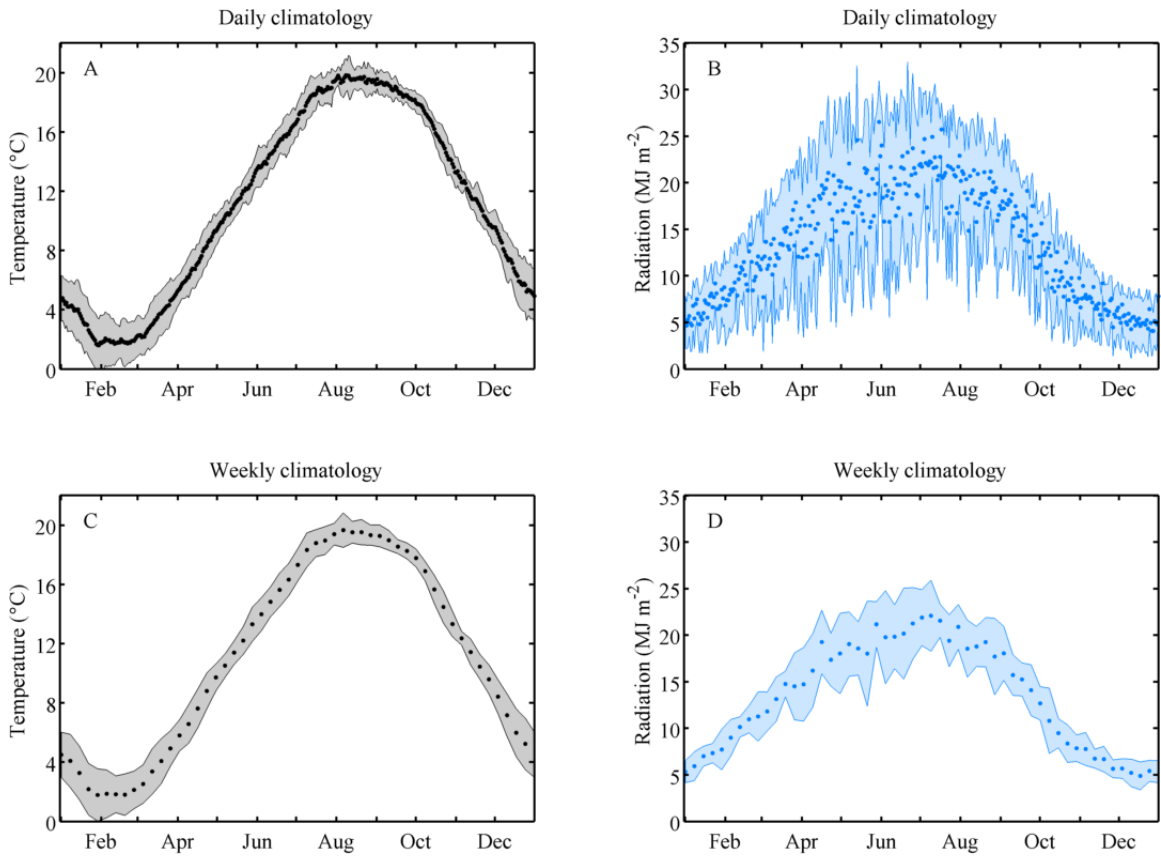
Figure 3-1: Daily climatologies at MVCO. A. Temperature (°C). B. Solar radiation (MJ m^{-2}). Weekly climatologies. C. Temperature (°C). D. Solar radiation (MJ m^{-2}).
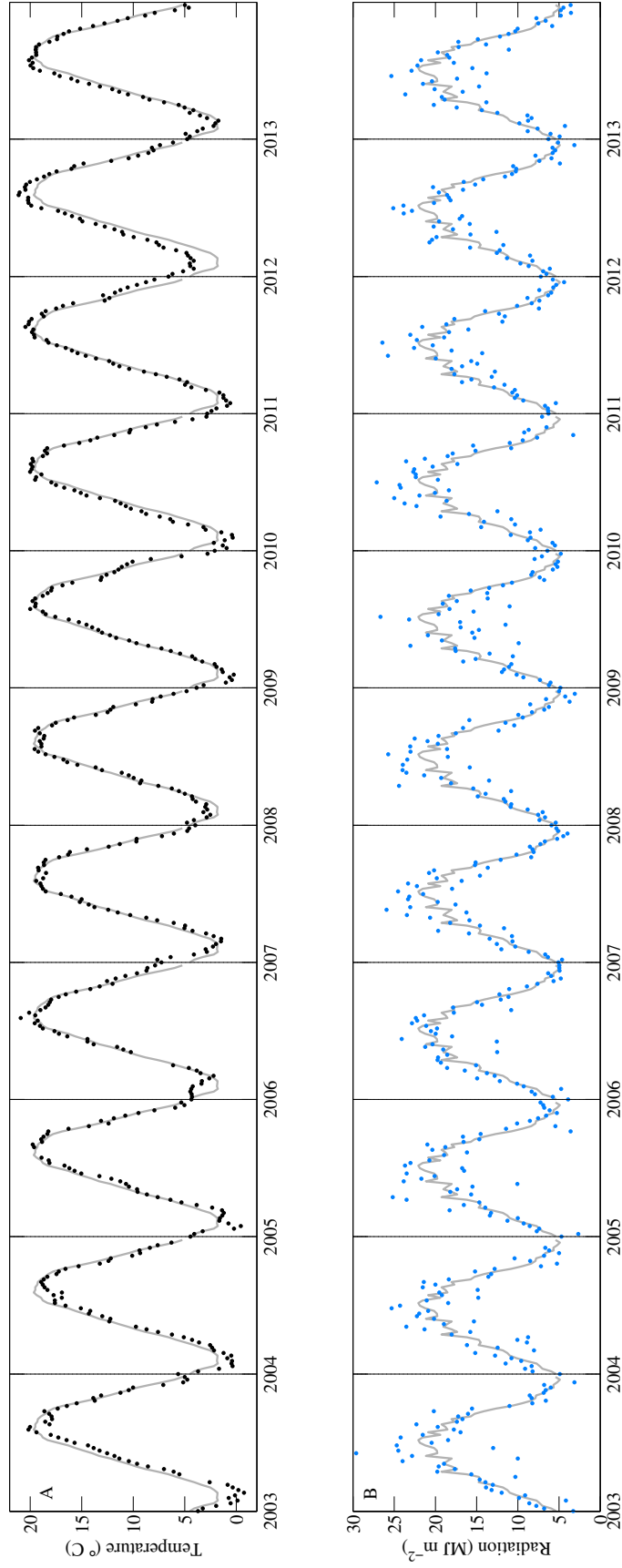
Figure 3-2: A. Weekly averaged records for all years in dataset. A. Temperature B. Solar radiation. Gray lines in each panel are the respective weekly climatologies for comparison.
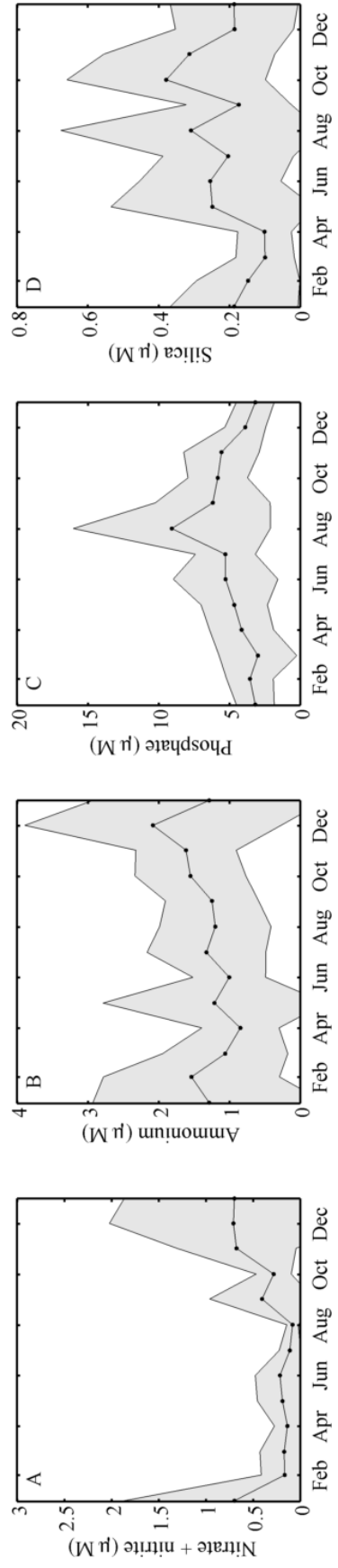
Figure 3-3: Monthly climatologies for nutrient concentrations (μM). A. Nitrate + nitrite. B. Ammonium. C. Inorganic phosphate. D. Silicate. Monthly climatologies were constructed with 181 data points.
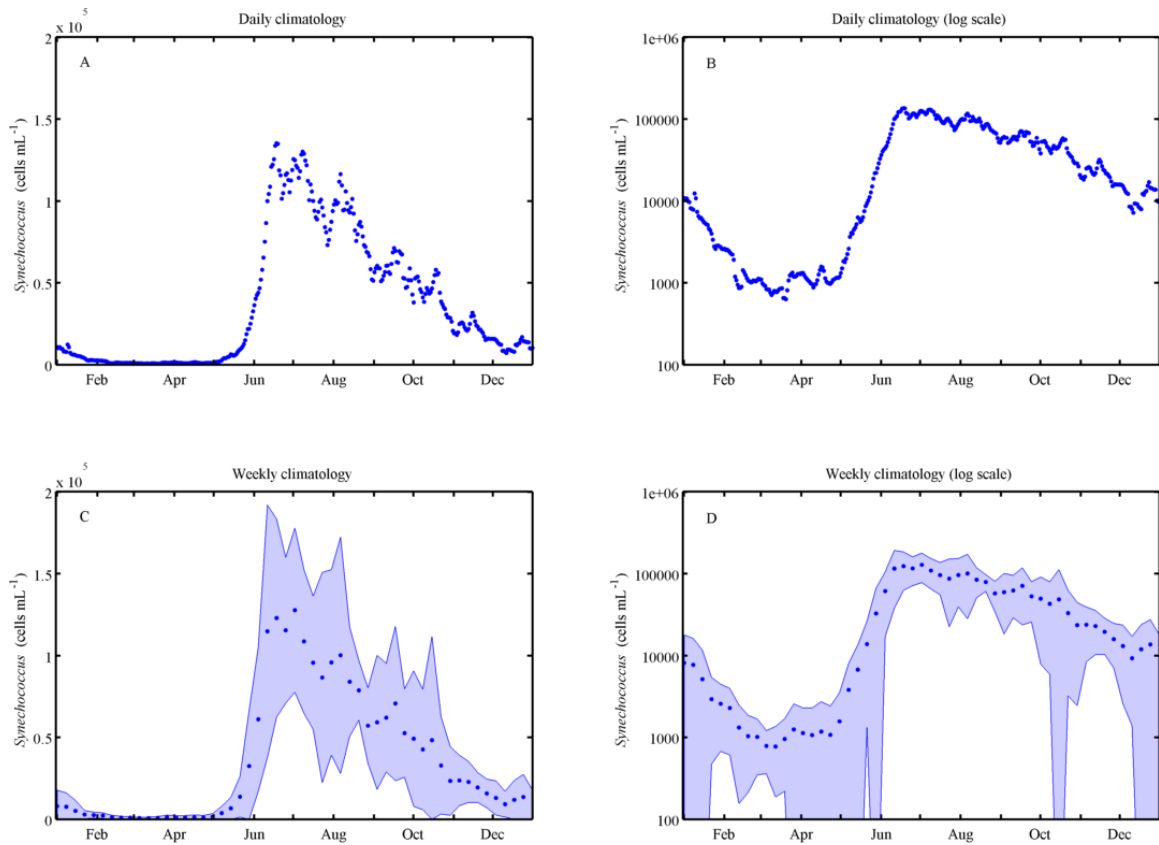
Figure 3-4: Daily climatologies. A. Cell abundance (cells mL^{-1}). B. Cell abundance presented on a log scale to visualize lower abundance patterns. Weekly climatologies of cell abundance. An average for each week was computed before computing the weekly climatology. C. Cell abundance D. Cell abundance presented on a log scale. Shaded region indicates one standard deviation away from mean.
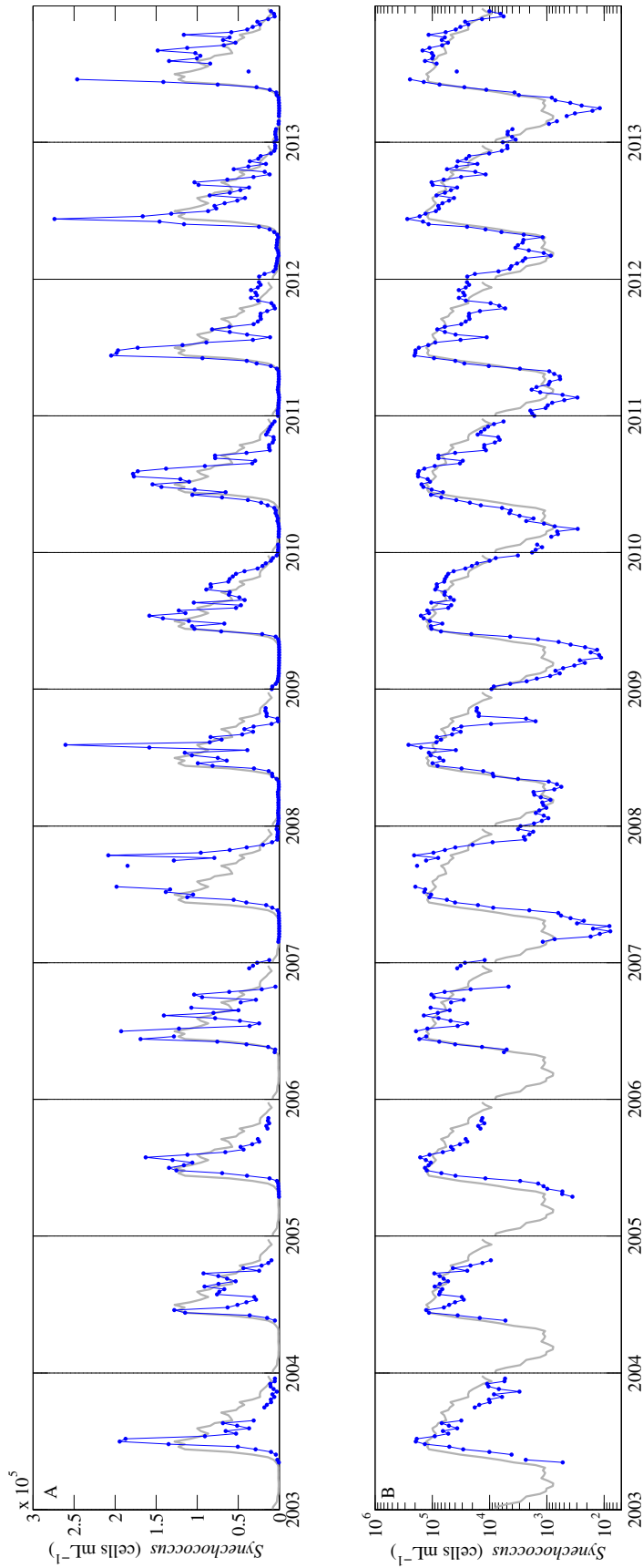
Figure 3-5: A. Weekly averaged *Synechococcus* cell concentration for each year. B. Weekly averaged *Synechococcus* cell concentration, but displayed on a log scale to show variability at low cell concentrations. Gray lines in each panel denote weekly climatology for comparisons.
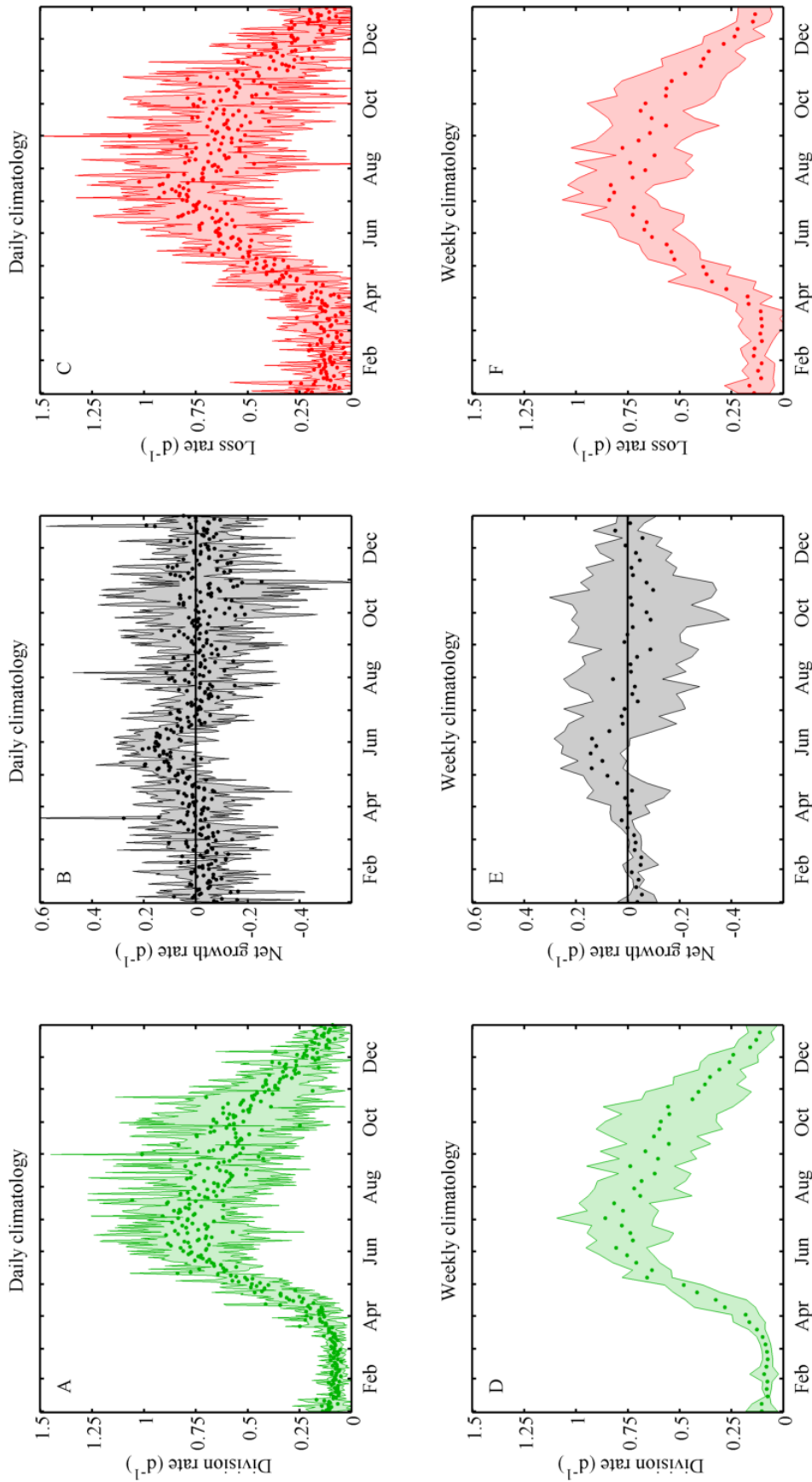
Figure 3-6: Daily climatologies. A. Division rate (d^{-1}). B. Net growth rate (d^{-1}). C. Loss rate (d^{-1}). Weekly climatologies. D. Division rate. E. Net growth rate. F. Loss rate. Shaded regions indicate values within one standard deviation of the mean.
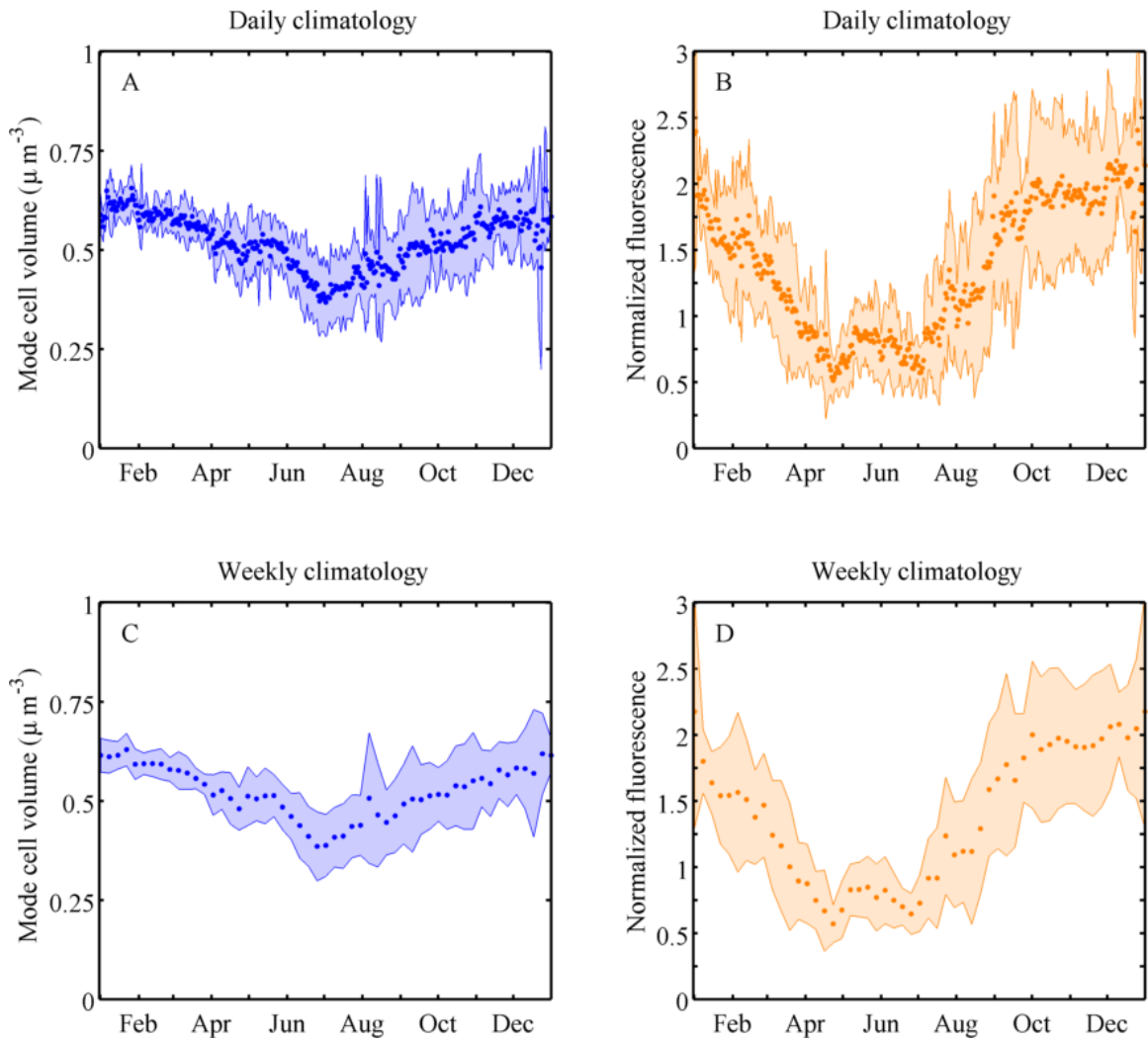
Figure 3-7: Daily climatologies. A. Mode *Synechococcus* cell volume (μm^3). B. *Synechococcus* bead-normalized cellular fluorescence. Weekly climatologies. C. Mode *Synechococcus* cell volume. D. *Synechococcus* bead-normalized fluorescence. Shaded regions indicate one standard deviation away from mean.
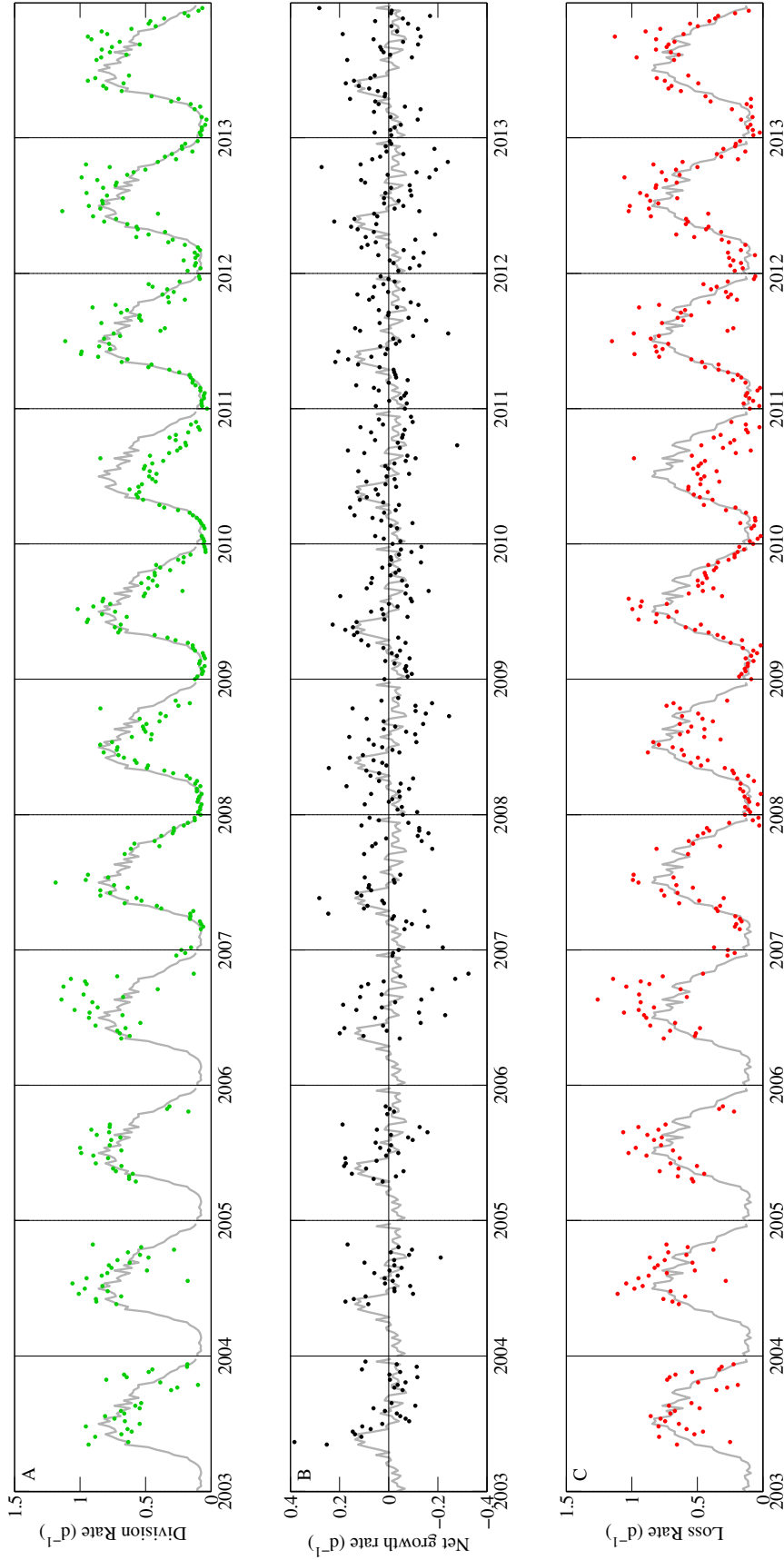
Figure 3-8: Time series of weekly averaged rates. A. Division rate (green). B. Net growth rate (black) C. Loss rate (red). Gray lines in each panel are the respective weekly climatologies for comparison.
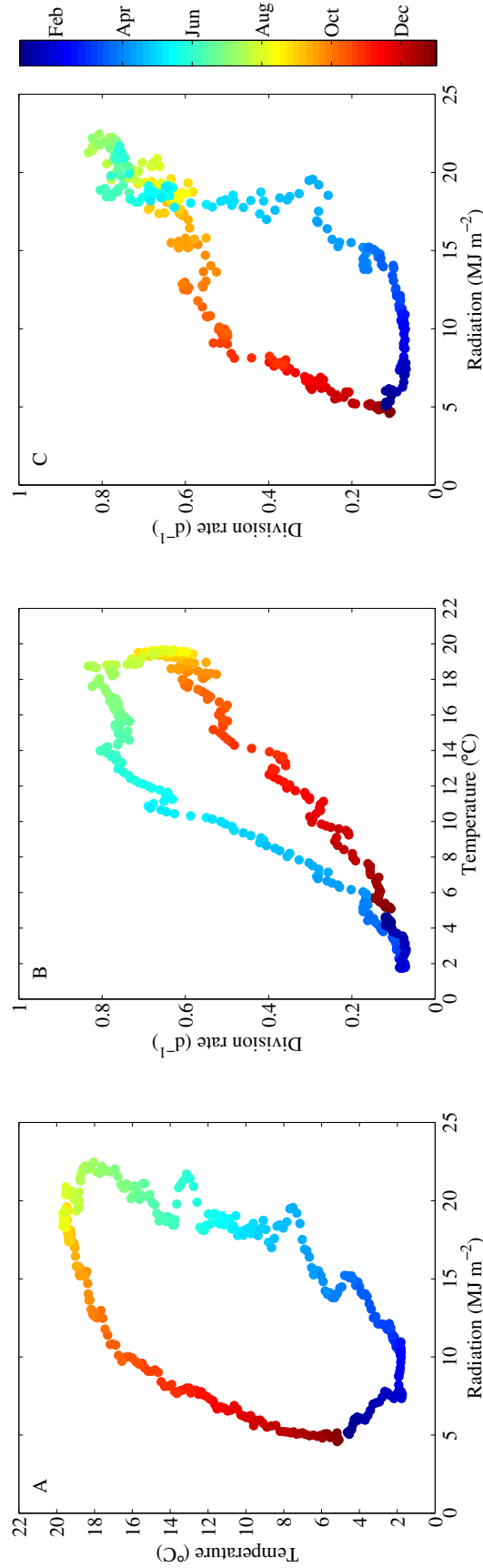
Figure 3-9: Relationships between daily climatologies. A. Solar radiation vs. temperature. B. Division rate vs. temperature. C. Division rate vs. solar radiation. All climatologies were smoothed with a 7-day running average. Symbol color denotes time of year.
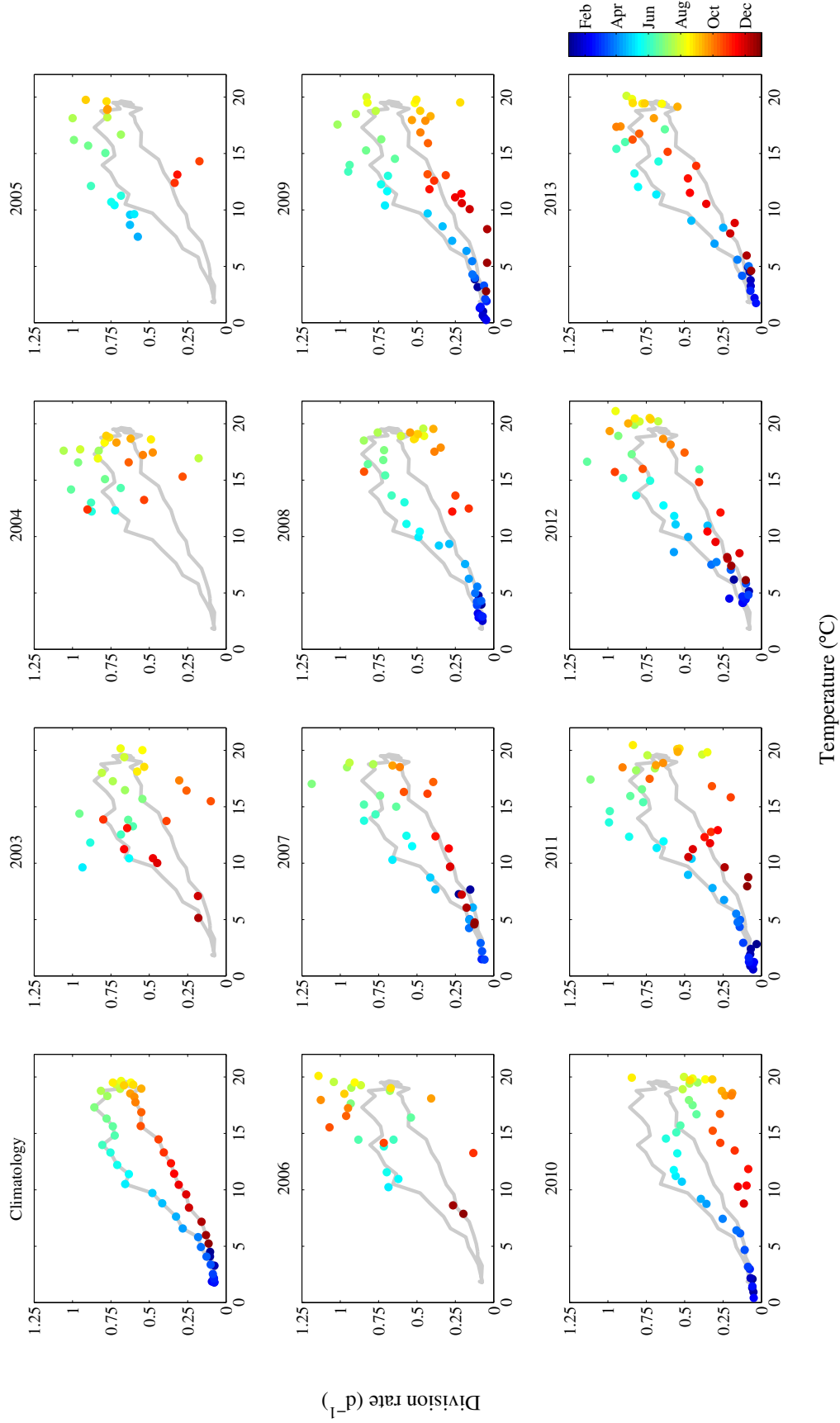
Figure 3-10: Relationship between weekly averaged division rate and weekly averaged temperature for each year in the time series. First panel shows the relationship between the weekly climatology for division rate and temperature. Color indicates time of year. Climatology reproduced on each panel for reference.

76
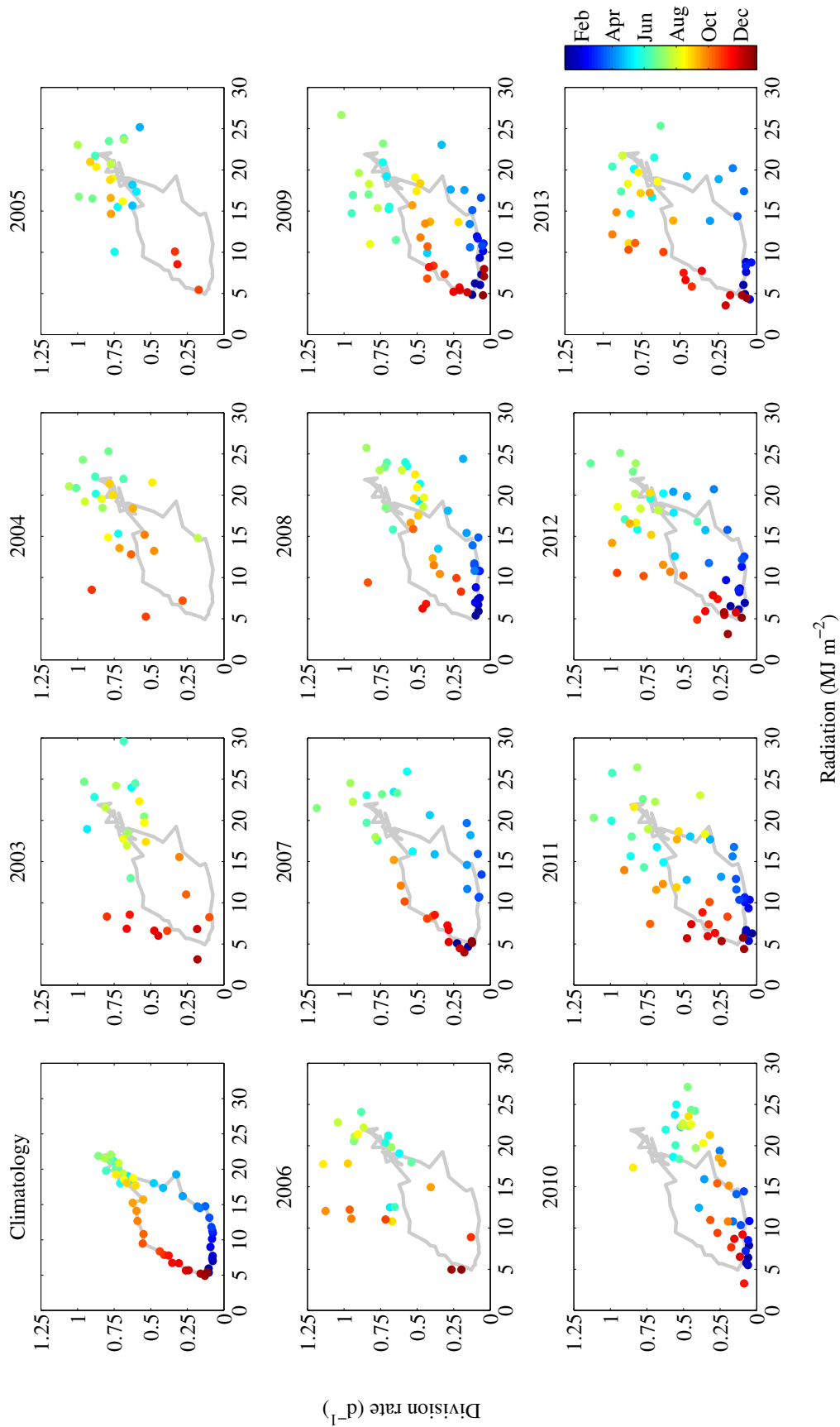
Figure 3-11: Relationship between weekly averaged division rate and weekly averaged solar radiation for each year in the time series. First panel shows the relationship between the weekly climatology for division rate and solar radiation. Color indicates time of year. Climatology reproduced on each panel for reference.
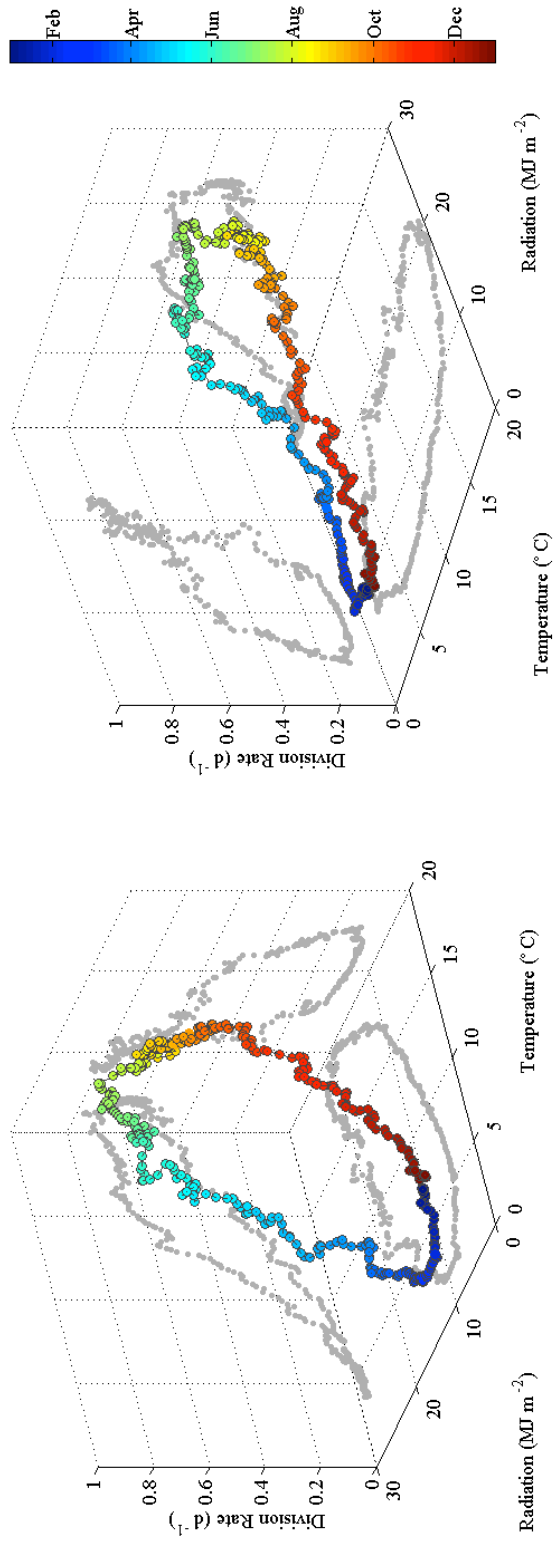
Figure 3-12: Three dimensional view of smoothed daily division rate against temperature and light. B. Same as in panel A, but shown from a different angle. Panels contain same climatological data as in 3-9.
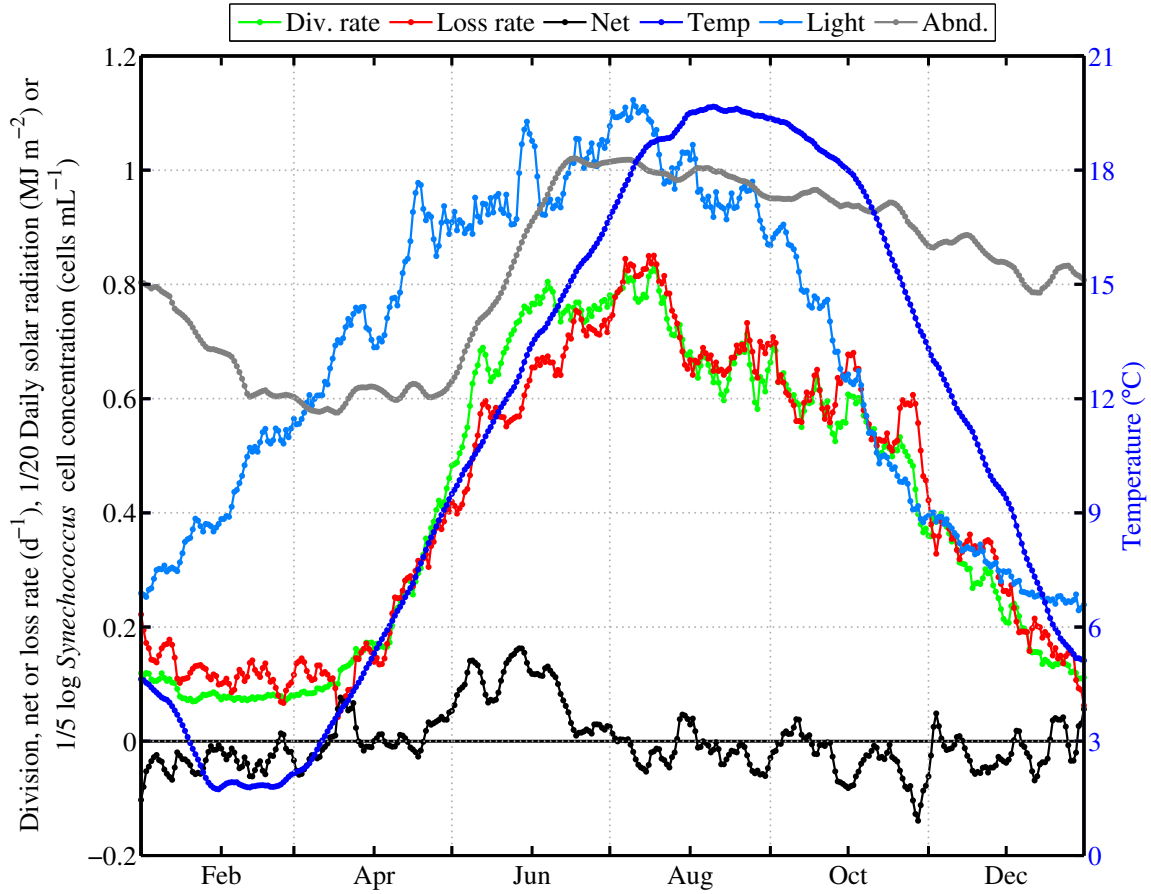
Figure 3-13: Smoothed (7-day running mean) daily climatologies of net growth rate (d^{-1}), division rate (d^{-1}), calculated loss rate (d^{-1}), temperature $(^\circ C)$, scaled solar radiation $(MJ\ m^{-2})$ and scaled log *Synechococcus* cell abundance.
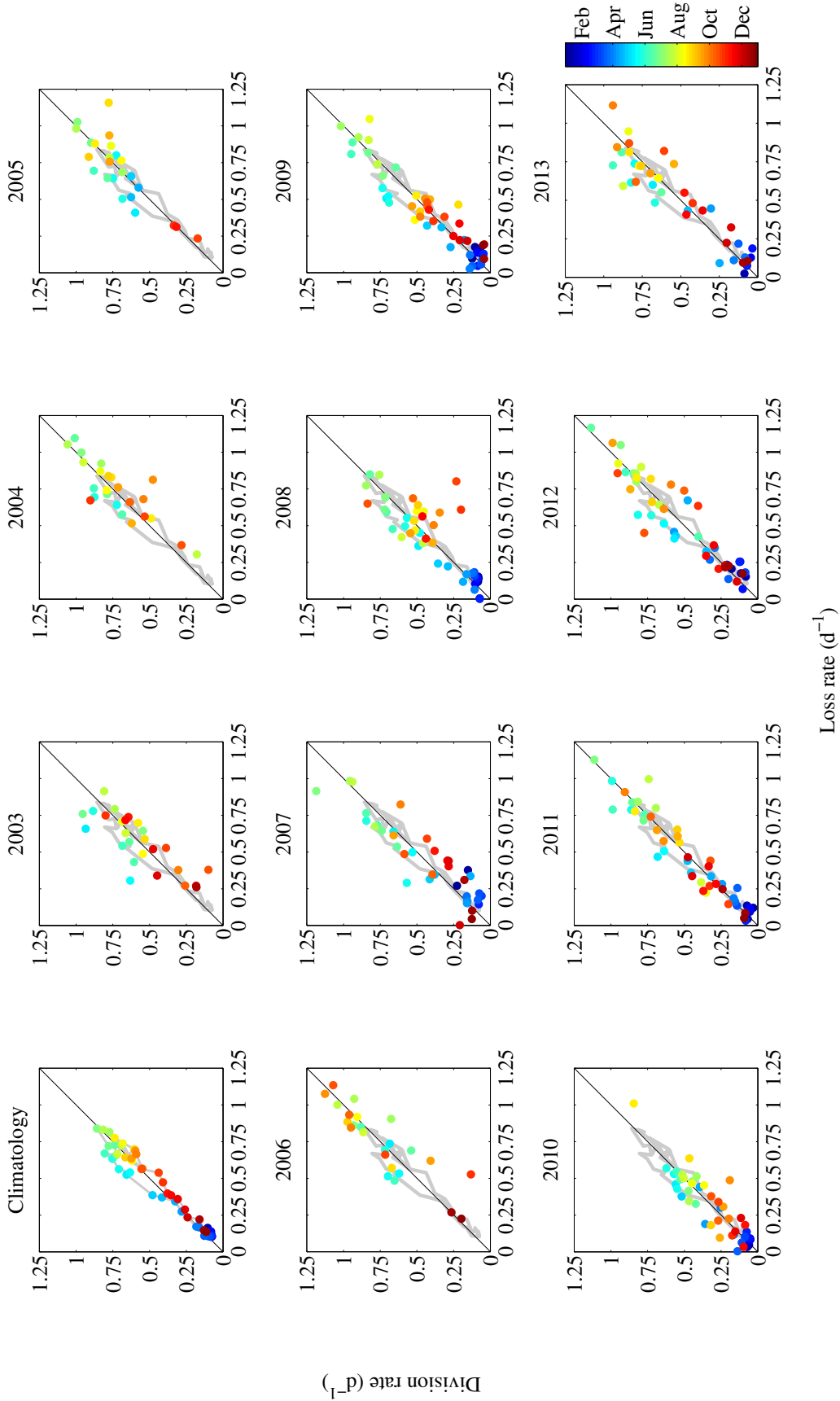
Figure 3-14: Relationship between weekly averaged division rate and weekly averaged loss rate for each year in the time series. First panel shows the relationship between the weekly climatology for division rate and loss rate. Color indicates time of year. Climatology reproduced on each panel for reference.
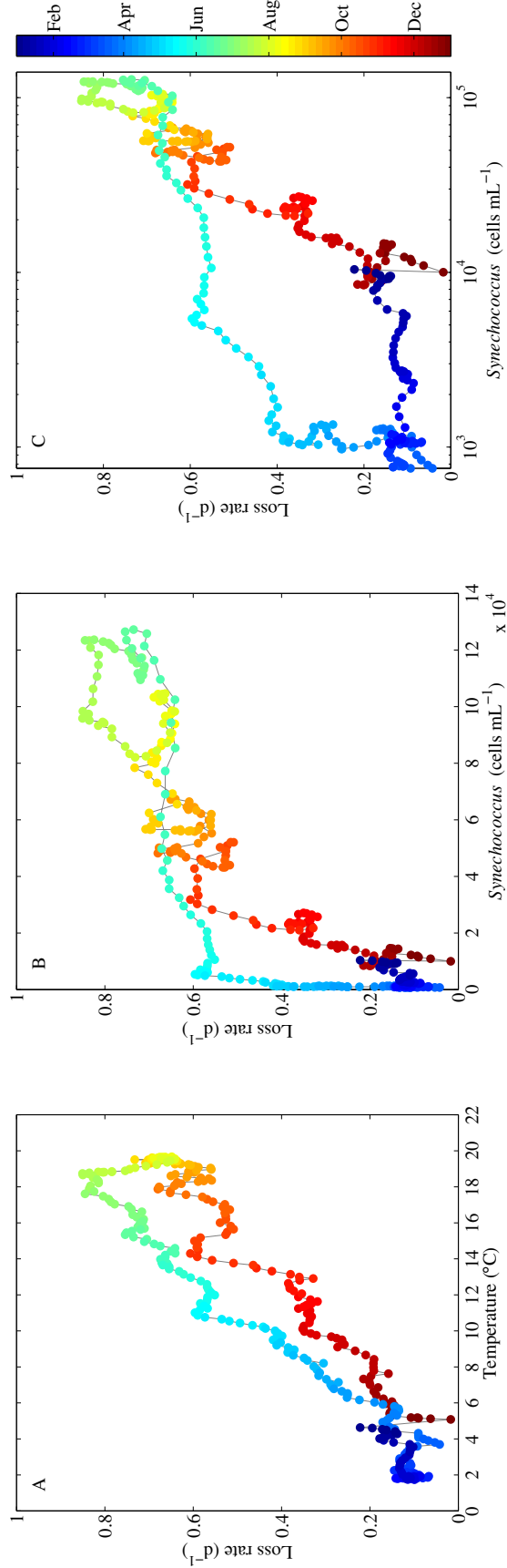
Figure 3-15: Relationships between smoothed daily climatologies. A. Loss rate vs. temperature. B. Loss rate vs. *Synechococcus* cell concentration. C. Same as in B, but with cell concentration shown on a log scale. Color indicates time of year.
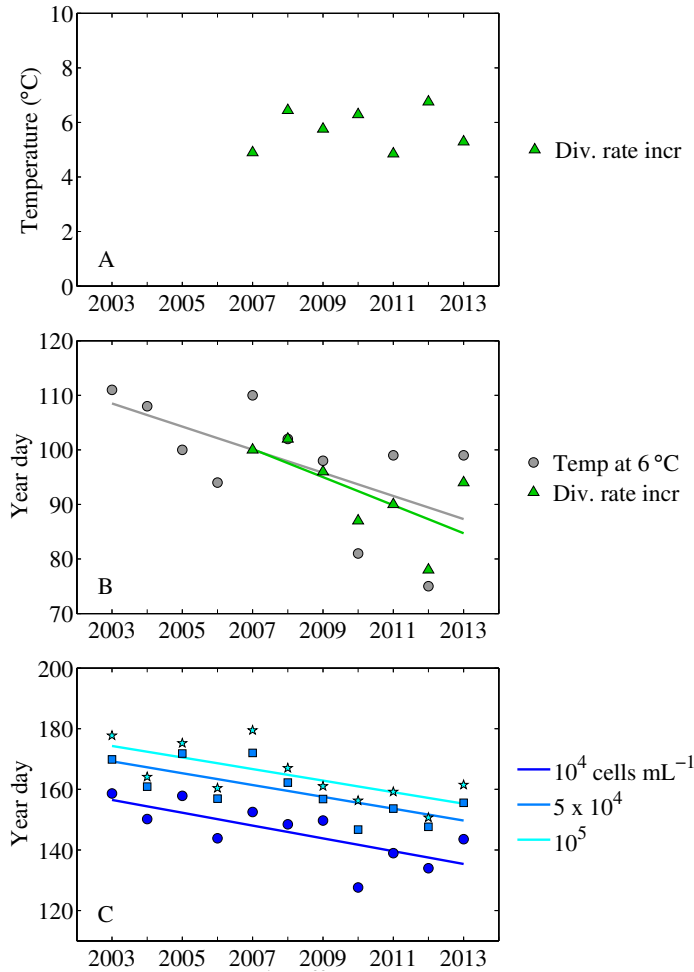
Figure 3-16: A. Temperature at the time when division rate began increasing for each year when observations of the entire spring bloom were available. B. Year day at which temperature (grey circles) reached 6°C for each year and the year day at which division rate began to increase (green triangles). Lines are linear regressions of year day for temperature increase (grey line) and year or division rate increase (green line) and year. C. Year day at which *Synechococcus* abundance reached different cell concentrations for each year as indicators of bloom progress: 10^4 cells mL^{-1} (dark blue circles), $5 \cdot 10^4$ cells mL^{-1} (light blue squares), 10^5 cells mL^{-1} (cyan stars).Lines are linear regressions between year day and year for each cell concentration (indicated by same color as marker symbol).
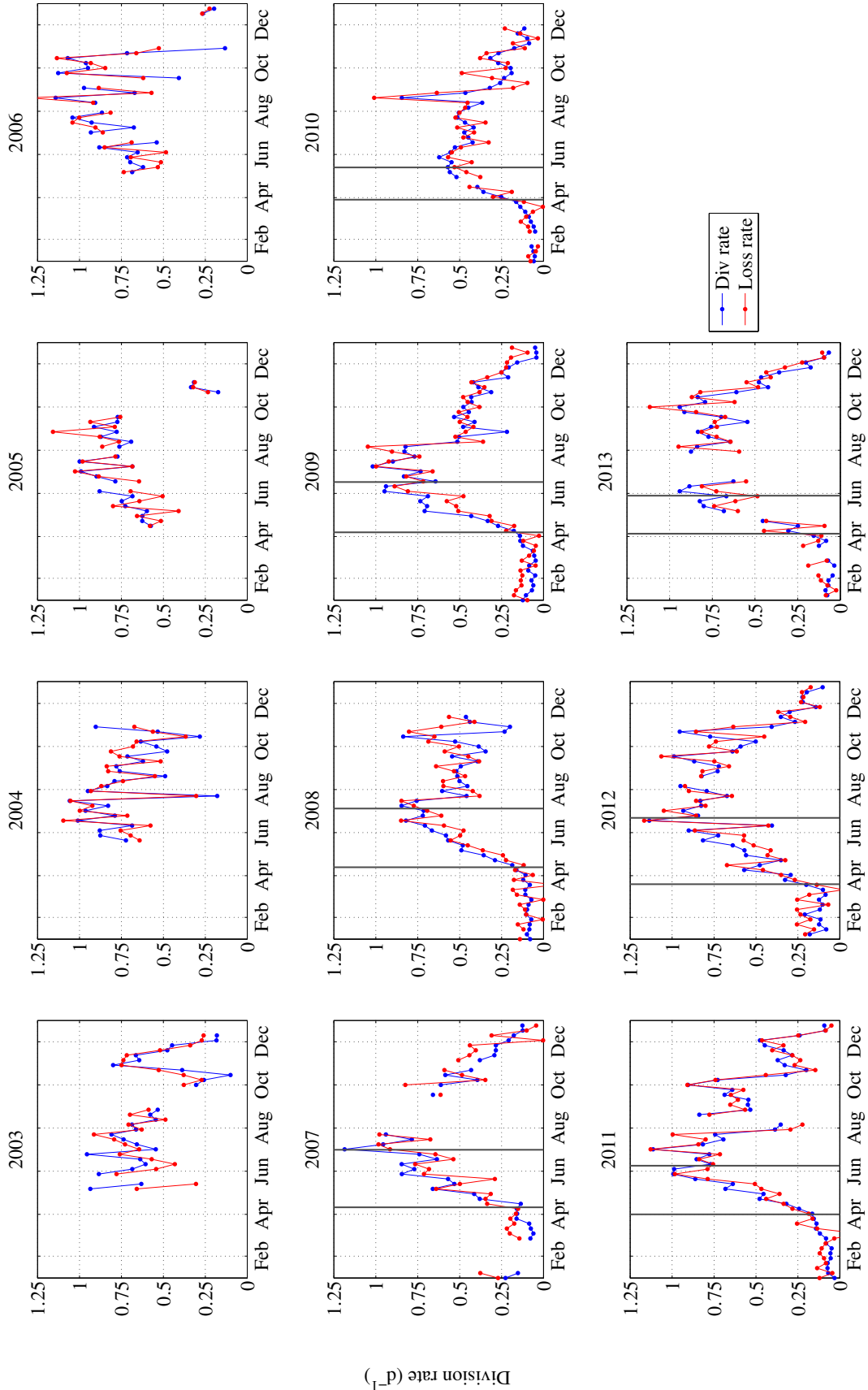
Figure 3-17: Weekly binned division rate (blue line) and loss rate (red line) for each year from 2003-2013. Solid vertical lines indicate the points identified as the start and end of the spring division rate increase.
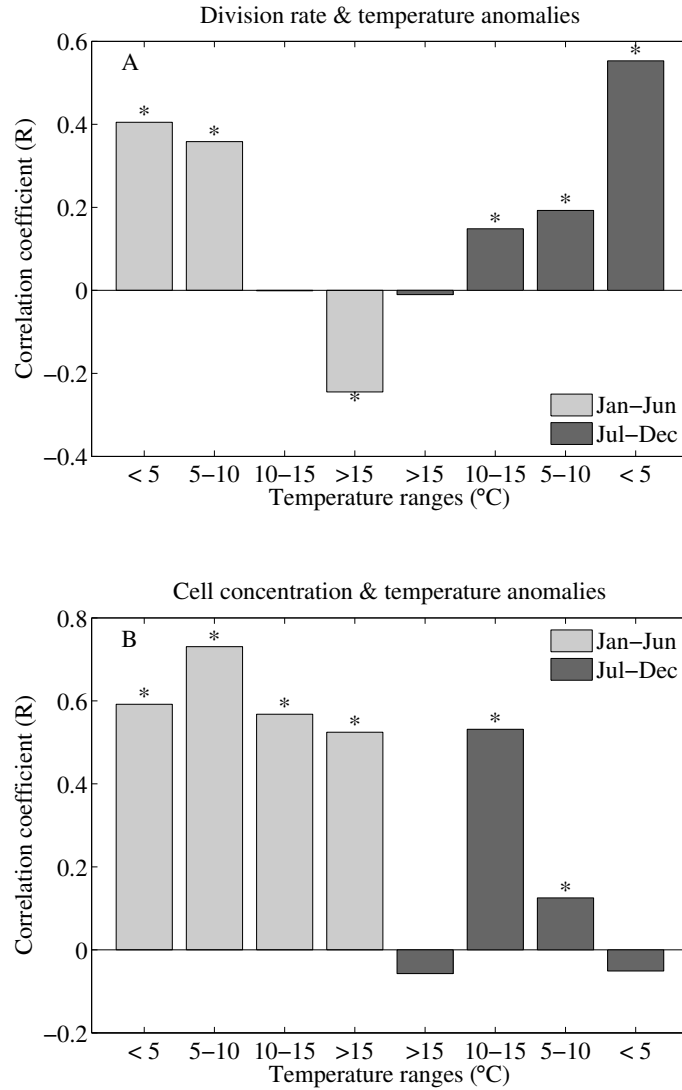
Figure 3-18: A. Value of correlation coefficient (R) between division rate anomaly and temperature anomaly for the first and second half of the year (light grey and dark grey bars, respectively) for 4 different temperature intervals. B. Value of correlation coefficient (R) between cell concentration anomaly and temperature anomaly for the first and second half of the year (light grey and dark grey bars, respectively) for 4 different temperature intervals. Asterisks above bars indicate that the correlation was significant ($p < 0.05$).
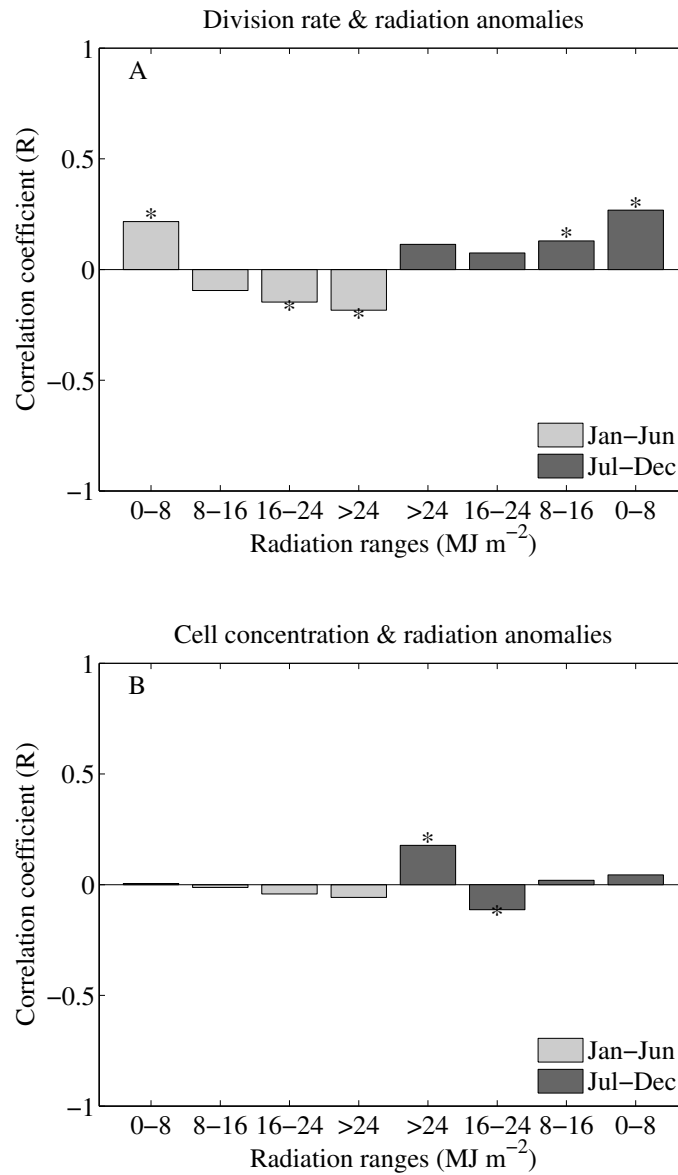
Figure 3-19: A. Value of correlation coefficient (R) between division rate anomaly and solar radiation anomaly for the first and second half of the year (light grey and dark grey bars, respectively) for 4 different solar radiation intervals (in MJ m^{-2}). B. Value of correlation coefficient (R) between cell concentration anomaly and solar radiation anomaly for the first and second half of the year (light grey and dark grey bars, respectively) for 4 different solar radiation intervals (in MJ m^{-2}). Asterisks above bars indicate that the correlation was significant ($p < 0.05$).
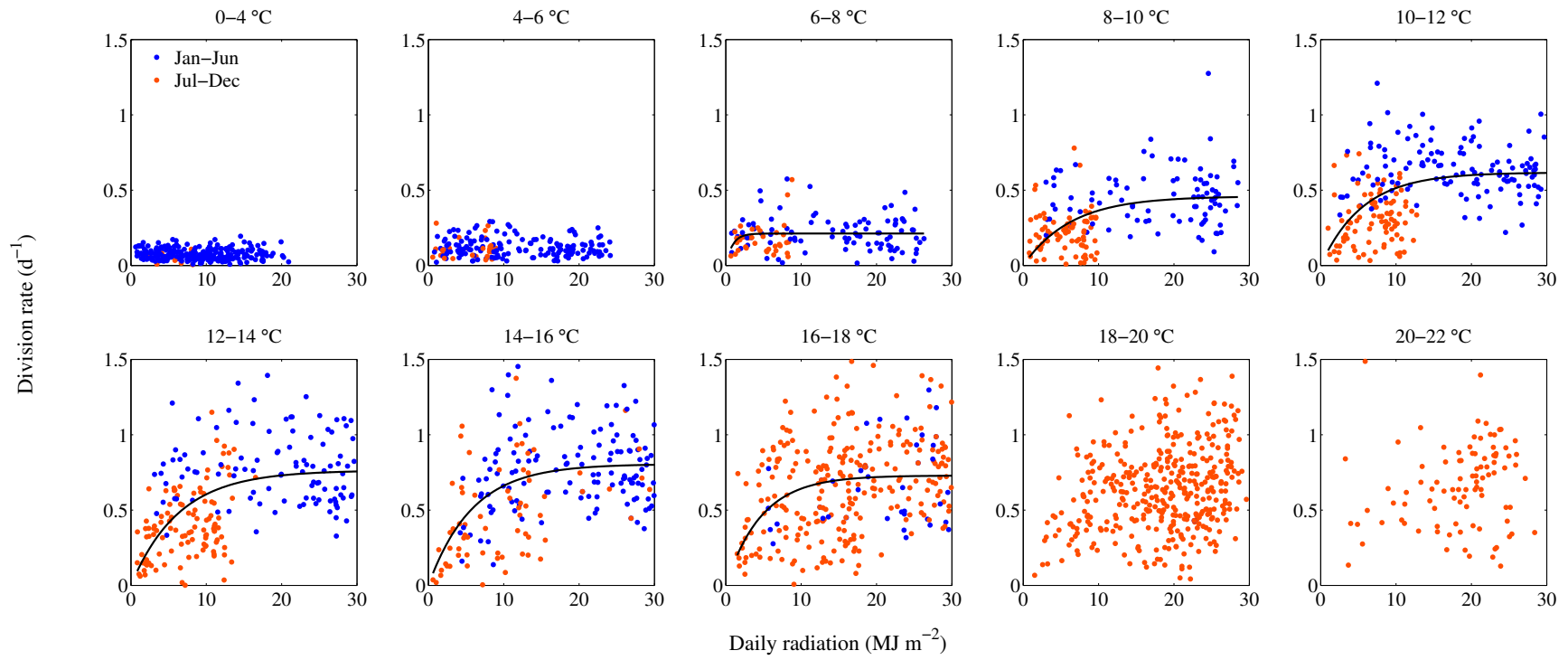
Figure 3-20: Relationships between daily division rate and solar radiation values for different 2-degree temperature intervals. Blue markers indicate values from January-June and dark orange markers indicate values from July-December. Black curves are minimized least square fits to the function: $\mu = \mu_{max}\left(1 - \exp\left(\frac{-\alpha E}{\mu_{max}}\right)\right)$, where μ is division rate, E is daily radiation, and α and μ_{max} are fitted parameters.

Chapter 4

Diversity of *Synechococcus* at the Martha's Vineyard Coastal Observatory: Insights from culture isolations, clone libraries and flow cytometry

4.1 Introduction

The cyanobacterium *Synechococcus* is a globally important primary producer in the world's oceans. This picophytoplankter (~ 1 μm diameter) is responsible for a substantial amount of carbon fixation (up to 20% in coastal systems, [59, 43]). Thus it is important that we understand the factors that affect *Synechococcus* abundance and enable it to be ecologically important under a wide range of environmental conditions. One of these key factors appears to be the high level of diversity contained within the *Synechococcus* genus. Studies of molecular phylogeny have resolved isolated strains and environmental sequences into a total of 20 well-defined clades distributed over three main sublcusters (5.1, 5.2 and 5.3) [24, 92]. These clades have been supported by phylogenies constructed from a variety of loci, including *rpoC1* [79, 106], ITS [3, 37], *narB* [71, 76, 77], *ntcA* [82, 84] and *petB* [64]. Clade designation was recently shown to be congruent across these genetic markers [4] and multi-locus sequence analysis of core genes provides evidence that the clades are in fact distinct lineages [64].

This genetic diversity may be representative of physiological or ecological diver-

sity, such that each clade (or closely related clades) corresponds to an ecotype that occupies a distinct niche [3]. This relationship between genetic diversity and ecological physiology has been well documented in the sister genus, *Prochlorococcus*, where genetically distinct clades show differences in light acclimation and nutrient utilization [69, 66, 68, 87]. Differences in clade physiology explain vertical distributions of *Prochlorococcus* in the water column, as well as clade biogeography across ocean basins [116, 5, 47, 120].

Similar relationships between genetic designation and physiological characteristics have been shown for some clades of *Synechococcus*. For example, strains belonging to clade III exhibit a motility that is unique to this clade [115, 108]. Clades also exhibit differences in nitrogen (N) utilization; some clades are unable to grow on or demonstrate reduced growth rate with nitrate (clade CRD2 and XV, respectively; [68, 3]), while others are able to utilize different N sources, such as urea and amino acids [68]. The response of growth rate to temperature can differ among clades as well as the response to temperature stress [83]. Clades also exhibit differences in light harvesting pigments [114, 3, 97, 35, 18], such that wavelength partitioning of available light is thought to be one axis of ecotype differentiation [101, 34]. Many members of subcluster 5.1 contain phycoerythrin (PE) in their pigment complements, and some of these clades can chromatically adapt to different light environments (clades I, III, XV and XVI; [80, 3]). Many members of subcluster 5.2 contain only phycocyanin (PC) as their light harvesting phycobilipigment (type-I pigment, [92]). These pigment distinctions are not absolute, however, as some strains from subclusters 5.2 and 5.3 contain PE, and some PC-only strains have been found in subcluster 5.1 (clade VIII). Six et al. [97] suggest that phycobilisome rod genes have evolved independently from the rest of the core genome, and are likely to have undergone horizontal gene transfer.

These physiological differences so far have not been sufficient to explain observed clade distributions in the ocean. Clades often co-occur [28, 17, 84], with reports of as many as 6 clades found at once [4]. Nonetheless, biogeographical and time series studies have begun to identify environmental factors that may shape some clade distributions. For instance, clades I and IV are typically found in colder, nutrient-rich, coastal waters at latitudes greater than 30° N and 30° S, and clade members CB4 and CB5 are also found in coastal waters and in estuaries [17, 12, 37]. Other clades seem to prefer warmer and more oligotrophic environments; clades II and III are typically found in tropical waters, with clade II having a much wider distribution into subtropical areas [120].

A complex set of interacting factors likely determine clade distributions. Abun-

dance will be governed by bottom-up conditions, such as light, nutrient availability, and temperature; top-down factors, such as grazing and viral lysis [71, 121, 6]; as well as side-ways interactions, such as with heterotrophic bacteria [95]. These factors change over different time scales, such as across seasons and over water columns with variable mixing. The time scales of environmental changes may contribute to the ability of clades to simultaneously coexist. Consistent with this idea, time series studies of clade abundances have demonstrated shifting dominance. In California coastal waters, Tai and Palenik [106] found that clades I and IV were always dominant, but with changing relative abundance over the seasonal cycle, while clades II and III only appeared during autumn and even then at relatively low abundance. In the Gulf of Aqaba, Post et al. [84] observed a succession of clades across the transition from winter mixing to summer stratification, which led to insights of possible preferred nutrient environments for clades I, III, and V/XII. These studies highlight the need to understand how temporal aspects of the environment shape and maintain *Synechococcus* diversity. At present, there is little knowledge of how seasonal environmental changes affect clade abundances in North Atlantic coastal waters, including on the New England Shelf. To better understand how clade patterns may change over time, we used *ntcA* clone libraries and culture isolations to assess the diversity of *Synechococcus* at the Martha's Vineyard Coastal Observatory (MVCO) over a total of three annual cycles. We further investigated the abundance of both PE-containing and PC-only *Synechococcus* in these coastal waters by analyzing time series samples with a flow cytometric setup that separates PC-only *Synechococcus* from picoeukaryotes.

4.2 Materials and Methods

4.2.1 Sample collection

Seawater samples were collected near the MVCO offshore tower (41° 19.500' N, 70°34.0' W) or at the offshore node (41° 20.195'N 70° 33.3865' W); ~3 km from south shore of Martha's Vineyard, (Fig. 4-1) at roughly bimonthly-to-monthly intervals over a 3-year period from 2010-2012 (43 total samples). Seawater was collected at 2, 6, 10 and 15 m depth with Niskin bottles attached to a rosette sampler or at the surface via bucket sample. Glutaraldehyde was added to a 5-mL aliquot of the water sample to a final concentration of 0.1% by volume for later flow cytometric analysis. These samples were incubated for 10 min at room temperature before being

frozen in liquid N_2. Samples for nutrient analysis were immediately filtered through a 0.2 μm Sterivex® filter into acid-washed vials and frozen at -20 °C. Samples were analyzed for phosphate, ammonium, silicate, and combined nitrate + nitrite by standard autoanalyzer techniques at the Woods Hole Oceanographic Institution Nutrient Analytical Facility (Woods Hole, MA). Only surface water samples were used for subsequent DNA extraction and culture enrichments (see below). Near continuous measurements of water temperature and salinity were obtained with a MicroCat CTD (SeaBird Electronics) deployed on the MVCO offshore tower at 4 m below mean water level. When there were short gaps in this data record, MicroCat data from the MVCO offshore node (12 m depth) were substituted. These records have been previously shown to be very similar [81].

4.2.2   Flow cytometry analysis

A modified Epics V flow cytometer (FCM; Coulter Electronics Corp.) interfaced with a Cicero acquisition system (Cytomation, Inc.) was used to analyze preserved water samples. The instrument was equipped with a 5-W argon ion laser (Coherent Innova 90-5), and photomultipliers for 3 wavelengths of fluorescence detection and forward light scattering. Excitation was at 515 nm (300 mW) and a 540 long-pass barrier filter was used to eliminate scattered laser light from the fluorescence detectors. Fluorescent emissions were split by successive dichroic mirrors and interference filters to measure wavelength bands of 562-588 nm (PE fluorescence), 610-660 nm (PC fluorescence), and 660-700 (chlorophyll fluorescence) (see Fig. 4-2 for schematic and filters used). Forward light scattering was measured at $\sim$3-19° above the axis of the laser beam. Samples were allowed to thaw in water before analysis and were injected into the sheath flow (MilliQ water, Millipore) by a peristaltic pump (Harvard Apparatus) at 0.1 mL min^{-1}. Polystyrene microspheres (Polysciences Inc.) of diameter 0.5 μm (polychromatic) and 1.0 μm (red-fluorescing) were measured as reference particles.

PE-containing *Synechococcus* were determined from characteristic PE fluorescence values and forward light scattering [73]. PC-only *Synechococcus* were determined from values of PC fluorescence, PC-to-chlorophyll fluorescence ratio, and forward light scattering. These features allowed separation and enumeration of PC-only *Synechococcus* from picoeukaryotes. Values of these parameters fell within a well-confined range for cultures (Fig. 4-3), and these values were used to guide analysis of field samples. Specifically, an event was designated as a PC-only *Synechococcus* if it did not show any PE fluorescence, had minimum values of $1.5 \cdot 10^4$ arbitrary fluorescence units for

both PC and chlorophyll fluorescence, and had a PC:chlorophyll fluorescence ratio of > 0.5, but < 1.0. Cell concentration was determined from sample flow rate and analysis time.

4.2.3 *Synechococcus* isolation

On selected days (see markers in Fig. 4-5 and dates in Fig. 4-11), surface seawater was prefiltered over a 20 μm Nitex® mesh. Filtrate was then gravity filtered through either 1 or 2 μm polycarbonate filters (Poretics) to exclude larger cells. The final filtrate was amended with nutrients at one third the concentrations described for SNAX media [114]. Cycloheximide was added at a final concentration of 50 μg mL^{-1} to prevent growth of eukaryotic phytoplankton and nanoflagellate grazers. Tubes were incubated near ambient seawater temperature with light levels of 30-80 μmol quanta m^{-2} s^{-1}. After 4-6 weeks, enrichments were visually inspected before transfer into fresh SNAX media (full strength) and then routinely transferred approximately every 4 weeks. After 2-3 months, any enrichment incubating at or below 12 °C was moved to a higher temperature (15-18 °C). For selected tubes, two- to four-month old enrichments were plated out onto 0.8% agar SNAX plates with 2 mM NaSO$_3$; agar was cleaned as described by Waterbury et al. [114]. Plates were incubated for approximately 1 month until single colonies of *Synechococcus* could be identified either visually or by epifluorescence. Single colonies were picked and inoculated into liquid media. For longer-term culture maintenance, isolates were grown in SN media [114] or a variation of SN (half nutrient concentrations of SN with ammonium at 5x concentration in SNAX). Of these clonal isolates, we selected 143 to be identified by genotyping of *ntcA*, an N-regulatory gene. Isolates were selected on the basis of different pigment colors (peach, orange, golden, green, etc.) and colony morphologies (size of colony, raised, sunken, circular or polygons, etc.).

4.2.4 Spectral analysis

To characterize the pigmentation types of selected cultures, in vivo fluorescence excitation and emission spectra were obtained for cell suspensions with a SpectraMax M3 (Molecular Devices) spectrofluorometer. For the excitation spectra, emission at 680 nm was measured over the range 400-660 nm at 5 nm increments. For the emission spectra, excitation was at 515 nm and emission measured from 530-700 nm at 2 nm increments, with a 530 long pass cutoff. A culture was determined to contain PE if the excitation spectra exhibited a characteristic peak between 530-580 nm and

the emission spectra also had a peak in this region [117]. Cultures were determined to contain only PC if they did not show a peak representative of PE, but rather a peak between 600-650 nm, characteristic of PC. Only a subset of the entire culture collection was characterized; strains that were not analyzed were assumed to have the same pigment type of strains with the same culture color and clade type (i.e., if a strain color was peach and belonged to clade I, it was classified as containing PE).

4.2.5 Environmental sample DNA extraction and *ntcA* PCR amplification

On selected days (dates marked on Fig. 4-5 and Fig. 4-10), 2-3 L of surface seawater was prefiltered through a 20 μm Nitex® mesh and then filtered onto 0.2 μm Sterivex® cartridge filters (Millipore) under vacuum pressure of no more than 40 kPa. Approximately 1.8 mL of DNA cell lysis buffer (Qiagen) was added to each cartridge before freezing and storage at -80°C. For DNA extraction, samples were thawed on ice and to break open cells, approximately 200 μL of 0.5 mm zirconia-silica beads (BioSpec Products) were added to the cartridges, which were then shaken vigorously at 2500 rpm for 10 min. Continued DNA extraction was carried out with Qiagen Purgene reagents, but with the modified procedure described in Palacios et al. [78]. DNA concentration was determined with a NanoDrop 2000 Spectrophotometer (Thermo-Scientific). Depending on the sample, 120 to 1200 ng of DNA template was added to PCR reactions with the degenerate 1F/4R primer pair [82]. Final primer concentration was 2 μM in a total reaction volume of 50 μL with Qiagen Taq PCR Master Mix Kit reagents. BSA was added at 0.2 mg mL^{-1} final concentration. PCR reactions were performed on a GeneAmp PCR System 9700 thermocycler (Applied Biosystems), with an initial denaturation period of 4 min at 94°C; followed by 40 cycles of 1 min at 94°C, 1 min at 45°C, 1 min at 72°C; and then a final extension step at 72°C for 7 min.

4.2.6 Culture isolate DNA extraction and *ntcA* PCR amplification

Approximately 2 mL of dense culture isolate was centrifuged at 9300 x g for 6 minutes to pellet cells and DNA was extracted from this pellet with a Qiagen DNeasy Plant Kit, following manufacturer's instructions with the exception of final elution volume (75 μL). Approximately 10-30 ng of DNA was added to PCR reactions with 1AF

92

and 4AR primers (targeted *Synechococcus* primers, [82]). Final primer concentration was 2.5 μM in a total reaction volume of 50 μL with Qiagen Taq PCR Master Mix Kit reagents. BSA was added at 0.2 mg mL^{-1} final concentration. Reactions were preformed on a GeneAmp PCR System 9700 thermocycler (Applied Biosystems), with an initial denaturation period of 4 min at 94 °C; followed by 30 cycles of 1 min at 94 °C, 30 sec at 55 °C, 30 sec at 72 °C; and then a final extension step at 72 °C for 7 min.

4.2.7   *ntcA* clone libraries

All PCR products (an expected 449 bp fragment), from both the environmental and culture isolates, were gel purified with a Qiagen Qiaquick gel extraction kit. Cleaned products were cloned into TOPO vectors for sequencing (TOPO TA Kit, Invitrogen) and transformed into chemically competent *E. coli* TOP10 cells (Invitrogen) following manufacturer's instructions. For isolates, at least 5 positive colonies (determined by blue/white selection on X-Gal, kanamycin LB plates) were picked. Plasmids were obtained via automated plasmid purification with a BiomekFX at the Josephine Bay Paul Center Keck Facility (Marine Biological Laboratory, Woods Hole, MA). Sequencing reactions used BigDye Terminator chemistry (Applied Biosystems) and contained at least 200 ng of purified plasmid and M13 reverse primer (15 μM, from TOPO kit). Sequencing was performed at the Josephine Bay Paul Center Keck Facility with a 3730 DNA Analyzer (Applied Biosystems). Primer and vector sequences were removed and resulting sequences were identified by BLAST search against *ntcA* accessions in NCBI Genbank.

4.2.8   Phylogenetic analysis

Sequences identified as *ntcA* were aligned with the ClustalW algorithm in BioEdit (version 7.2.0, [33]). Operational taxonomic unit (OTU) construction and rarefaction analysis was carried out in mothur v.1.23.1 [94] with the furthest neighbor clustering algorithm. Distinct OTUs were designated at a 10% dissimilarity cutoff. Phylogenetic reconstructions were carried out in the ARB software package (version 5.3, Ludwig et al. [61]) with a maximum likelihood approach using RAxML [99] and a GTR GAMMA rate substitution model. Bootstrap analysis for support of tree branches was also carried out in ARB with rapid bootstrap analysis and 500 sample trees.

4.3 Results

4.3.1 Environmental conditions

Temperature at MVCO undergoes large seasonal changes from a minimum of around
0 °C up to a maximum of 22 °C during the three-year period of this study (Fig. 4-4A).
Salinity was typically within the range 31-32.5. The concentration of nitrate + nitrite
was usually below 1 μM, with the majority of the samples below 0.5 μM and often at
the limit of detection for the autoanalyzer technique (0.05 μM) (Fig. 4-4B). Higher
nitrate + nitrite concentrations (0.75-1 μM) occasionally occurred during fall. The
concentration of phosphate was also usually low (< 0.25 μM for 94% of samples).

4.3.2 Flow cytometry analysis

Consistent with our previous reports for this site [38], cell abundance of PE-containing
Synechococcus followed a repeatable seasonal pattern of low wintertime concentrations
of a few hundred cells mL^{-1} to greater than 10^5 cells mL^{-1} in summertime (Fig. 4-5).
Large chances in abundance (an order of magnitude) were observed during the late
summer and fall. By contrast, only a few samples over three seasonal cycles appeared
to contain signatures that matched our criteria for identification as a PC-only *Syne-
chococcus* (see Fig. 4-3 for an example). The maximum observed concentration was
roughly 3-fold less than that of concurrent PE-containing *Synechococcus*. The ma-
jority of small ($<\sim2$ μm), non-PE containing cells were classified as picoeukaryotes
(Fig. 4-3).

4.3.3 Spectral analysis

Isolates belonging to clades I, II, III, VI, VII, CB5, 5.2MV2, and 5.3I demonstrated
excitation and emission spectra consistent with the presence of PE. These isolates
exhibited peaks between 565-580 nm for emission spectra (Fig. 4-6A) and between
540-560 nm for excitation spectra (Fig. 4-6B). Isolates belonging to clades VI, CB5,
and 5.2MV2 had shifted excitation peaks (560-565 nm) relative to the rest of the
PE-containing isolates (peaks $\sim$ 540-545 nm). Isolates belonging to clades VIII,
CB4, 5.2MV1, and 5.2MV3 did not demonstrate characteristic excitation or emission
peaks of PE. Excitation peaks for these isolates were in the range 620-625 nm and
emission peaks 650-654 nm. These peaks are characteristic of PC. These different
spectra translated into a range of culture colors (Fig. 4-7). *Synechococcus* that were
determined to contain PE appeared as brown, dark red, pink, peach, or orange.

Synechococcus that only contained PC were dark bluish-green, green, or light yellow-green.

4.3.4 Diversity of environmental sequences

We obtained a total of 425 *ntcA* sequences from eleven amplicon libraries. Of these, 242 sequences were unique. The majority of the sequences ($\sim$97%; n = 414) belonged to clade I; ten sequences belonged to clade IV and only 1 belonged to clade CB4 (Fig. 4-8). There was considerable diversity within the clade I sequences, and OTU clustering separated these into 4 distinct subclades (Fig. 4-10, Fig. 4-9). Interesting, these subclades did not group with reference strains WH8020 and CC9311. Subclade IC sequences were found to dominate in all samples, ranging from 57% to 90% of the sequences in each clone library (Fig. 4-10). The second most abundant subclade was IB, which appeared to make up more of the population in spring and summer. Subclade IE sequences appeared in samples taken in late summer through early winter, and subclade IA only appeared in late fall 2010. The most diverse *Synechococcus* populations were observed in the late summer and early fall, and they included sequences from clade IV and clade CB4, as well as from subclade IA. However, members of these clades and subclade (IV, CB4, IA) appeared to be relatively rare in the environment. Two subclades were also identified within clade IV (IVA and IVB). While clone library results are not strictly quantitative, the relative frequency of each clade and subclade changed over the seasonal cycle in a manner that may reflect preferred environmental conditions.

4.3.5 Diversity of isolates

From the 17 enrichments, 143 isolates were identified from their *ntcA* gene sequences. Of these, 66 unique strains were identified (i.e., different *ntcA* sequences). Isolates with identical *ntcA* sequences did occur within the same enrichment, but also across different enrichments, with original sample collection dates sometimes separated by months (Table A.2). Analysis of *ntcA* phylogeny showed that the strains belonged to 12 different clades that spanned the three subclusters (Fig. 4-8). Isolates mapped to known clades I, II, III, VI, VII, and VIII of subcluster 5.1, three clades of subcluster 5.2, including CB4 and CB5, and one clade of sublcuster 5.3 (5.3I / X). Strains also clustered into two other clades, belonging to subcluster 5.2 that did not match to other known strains. These clades have been labeled as 5.2MV1 and 5.2MV2 until confirmation as either novel or known can be determined (strain representatives may

exist for which phylogenetic markers other than *ntcA* have been sequenced). Strains of clade 5.2MV1 contained PC as their primary pigment, whereas the one strain of clade 5.2MV2, MV1218, contained PE (see Table A.2). We have designated isolates that clustered closely with reference strain WH5701 as clade 5.2MV3.

Clade I isolates belonged to two of the subclades, IC and IE, observed from clone library sequences. Subclade IE contains reference strain WH8016. The other group, subclade IC, is not related to other known strains for which *ntcA* sequences are available (CC9311 and WH8020). Representatives of these subclades were consistently recovered in cultures from enrichments throughout the time series analysis (Fig. 4-11), which spanned a wide range of temperature and nutrient concentrations (Fig. 4-12).

Surprisingly, we also isolated strains considered more common to subtropical and tropical waters (i.e., clades II, III, and VII [120, 84]). Isolates of such clade members only occurred in late summer and early fall. Clade II types were isolated during September (2011 and 2012) and October (2012) and clade III was isolated during August and September (2012). Clade VI isolates were also only obtained in late summer and early fall and clade VII was only found during the fall. In general, isolations of clades II, III, VI, VII, CB5, and 5.3I tended to only occur when water temperature was > 16 °C and nitrate + nitrite concentration was relatively low (< 0.5 μM)(Fig. 4-12). The frequency of isolation of these clades was much lower than for clade I representatives (Fig. 4-12, Fig. 4-11). Clade occurrence did not correlate with phosphate concentration (data not shown).

We were able to culture PC-only pigment type strains from all but a few of the enrichments from MVCO (Fig. 4-11). Although not all of these isolates were sequenced (Fig. 4-11), those that were fell into clades VIII, CB4, 5.2MV1, and 5.2MV3 (Fig. 4-8). Isolation of these PC-only strains over the entire enrichment time series suggests that they were persistent members of the *Synechococcus* community with a year round presence at MVCO.

4.4 Discussion

The high degree of genotypic and functional diversity within the *Synechococcus* genus likely contributes to its importance in the worlds' oceans. *Synechococcus* and its sister genus *Prochlorococcus* offer excellent model systems in which to investigate how diversity is maintained and selected for in very closely related species. Studies that have investigated *Synechococcus* strain physiology and the biogeography and temporal aspects of clade distributions have enabled insights into the environmental

factors that determine their abundances. Despite such studies [120, 106, 84, 37], niche definitions for the different clades still remain elusive. Previous work has suggested that temporal changes in environmental conditions are important [106, 84]. Our studies of seasonal diversity patterns at MVCO support this view and also identify subclade level diversity as an important component of seasonality.

4.4.1 *Synechococcus* diversity at MVCO

Culture-dependent and culture-independent approaches illustrate that the *Synechococcus* assemblage found throughout the year at MVCO is diverse. We identified members of 13 different clades spanning all 3 known subclusters of marine *Synechococcus*, but members of clade I dominate the *Synechococcus* population over the entire year. This dominance is consistent with the known biogeography of clade I, which is primarily found in cooler, higher nutrient, coastal waters [120, 18, 106, 37], and has been shown to make up a majority of the *Synechococcus* assemblage in surrounding shelf areas [4]. While a preference for cooler water is likely to be a key factor explaining the presence of clade I at MVCO, the ability to survive colder winter water temperatures ($\sim$0-4 °C) may also contribute to clade I dominance at this site. Recently, Pittera et al. [83] demonstrated that clade I strains were able to grow at temperatures lower ($\sim$ 10-15 °C) than tolerated by strains from clades II and V. These authors also demonstrated that clade I strains were more tolerant of cold stress. At MVCO, strains of clade I were isolated from water at a range of temperatures (Fig. 4-12), but notably from water at $\sim$4°C, indicating that cells were still viable during this time. The PE-containing *Synechococcus* population reaches a minimum cell concentration of a few hundred cells per mL^{-1} during winter (Fig. 4-5), and the ability to survive these colder temperatures may be an important factor that allows this population to 'overwinter' until more favorable spring conditions.

Within clade I, we also found significant diversity at the subclade level, such that 4 different subclades could be resolved in the clone library sequences (Fig. 4-10, Fig. 4-9). Subclade IE is closely related to reference strain WH8016. Subclades IA, IB, and IC do not have known representatives in the *ntcA* tree and might be novel, but we cannot be certain they are not related to subclades previously observed with other markers. They could be related to the subclades observed by Tai and Palenik [106] in California coastal water or clade I strains analyzed by Mazard et al. [64]. Sequencing of the *rpoC1* or *petB* regions for these strains would be needed to resolve this. While not strictly quantitative, relative abundances of sequences in our

clone libraries suggest that not all of these subclades are equally represented when they co-occur, which could indicate that they have distinct environmental niches. While subclade IC appears to be dominant throughout the year, subclades IB and IE increased in relative abundance at different times of year (spring-summer, summer-early winter, respectively). Different subclade distributions have also been observed off the southern coast of California. Tai et al. [105] found that subclades within clades I and IV showed distinct depth distributions along a coastal to open ocean transect, suggesting that these subclades are likely to have preferred environments. At MVCO, we did not, however, find correlation between the relative abundance of sequences for each subclade and nutrient concentrations or temperature (data not shown). Isolations of subclades IC and IE usually occurred during the same conditions (although IE was isolated more frequently at warmer water temperatures). This suggests that other factors are important in regulating subclade abundances. As we now have two subclades in culture, going forward it will be possible to explore possible physiological and ecological differences that might govern this microdiversity.

Interestingly, clade IV, which has been reported to co-occur with clade I in other coastal waters [120], appears to have lower relative abundance at MVCO. Strain representatives were not able to be isolated and only $\sim 2\%$ of the clone library sequences belonged to this clade. While we cannot be sure that biases in clone libraries and culture isolations are not a factor, it may be that this clade is very rare at MVCO. This is consistent with observations of Ahlgren and Rocap [4] who found clade IV to be much less abundant than clade I at outer shelf locations 80-304 km south of MVCO. This is different from other coastal locations for which the abundance of clade IV usually matches or exceeds that of clade I [106]. The low abundance of clade IV at MVCO raises interesting questions as to the physiological differences between these two clades and why, for this coastal system, clade I is much more abundant. As highlighted by Ahlgren and Rocap [4], these questions are relevant for the wider shelf region of the northern Middle-Atlantic Bight.

Comparison of clade occurrence from culture enrichments with nutrient and temperature conditions at MVCO (at time of sampling) allows us to make possible inferences about the preferred environments of different clades. We were able to culture representatives of clades that have typically been found in either warmer or more oligotrophic environments (e.g., clades II, III, VII, 5.3I). Ahlgren and Rocap [4] found clades II and 5.3I at an outer shelf location (304 km south of MVCO) and detected clade III in samples from the Gulf Stream. One possibility for these clade occurrences at MVCO is that they are advected onto the inner New England shelf from more off-

shore locations and can only survive at this location during late summer and fall. The general circulation on the New England Shelf is part of the larger shelf circulation of the Middle Atlantic Bight, which is characterized by a southwestward along-shelf flow of relatively fresh water, with across-shelf offshore currents at the surface and bottom and onshore currents in the middle of the water column [57]. Shelf water is separated from saltier slope water by a shelf-slope front, but exchanges between these water types can occur due to frontal instabilities [31], eddies [30], warm-core ring shelf interactions [22, 48], and saline intrusions at the seasonal pycnocline [56]. Locally, there is a counter-clockwise recirculation just south of MVCO, which is strongest in the summer months [50]. Slope water intrusions plus this recirculation feature could make it possible for clades growing in warmer, saltier water to be advected and then retained near MVCO. During summer, water temperature is warm for this location ($\sim 20°$C) and nitrate levels are typically low or undetectable (Fig. 4-4), such that conditions may allow persistence of clades that would not typically thrive in coastal waters at other times of the year.

Clades VI, VIII, and CB5 were only isolated when water temperature was relatively warm. This is most apparent for Clade VI, which was cultured from seven different enrichments, but only when water temperature at time of sampling was > 16 °C. Previous studies have not provided unambiguous information on the global distributions of clade VI, as the probes used to date do not separate clades V, VI, and VII from one another [28]. Clade VI have been isolated from other coastal environments (Woods Hole Harbor, [114]; East Sea and East China Sea, [18]), suggesting tolerance or preference of coastal conditions.

4.4.2   PC-only *Synechococcus*

Strains of *Synechococcus* that only have PC as their light-harvesting pigment, have been isolated previously from either estuarine or near shore coastal waters [114, 17, 35, 18], and sequences that match these strain representatives have been found in similar regions [17, 12, 37]. PC-only *Synechococcus* are well suited to absorb the quality of light found in these more turbid waters [102, 103]. Given the near-shore location of MVCO, it is not surprising then that PC-only *Synechococcus* strains were isolated. However, consistently low to undetectable cell concentrations from flow cytometry analysis suggest that these strains may not be ecologically relevant at MVCO. These strains may have been transported from more estuarine sites and then grow poorly in the environmental conditions at MVCO. Although the site is exposed to the open

shelf, it is located on the inner shelf only 3 km from the south shore of the island of Martha's Vineyard. The sudden appearance and then rapid decline of PC-only *Synechococcus* in samples from October 2010 (Fig. 4-5, separated only by a week) is consistent with a hypothesis of a population advected to the site, but not able to persist.

The PC-only *Synechococcus* strains appeared to thrive, however, in the culture conditions, and in fact dominated many enrichments (Fig. 4-11). The media recipe used to culture and maintain the strains contained only 75% seawater. Many members of subcluster 5.2 are halotolerant, such that they do not require elevated salt requirements for growth [93]. If these strains grow better in lower salinity, then the enrichment conditions may have selected for these representatives. Anecdotal observations from our enrichments also suggest that these strains may persist at background levels for months in a low-nutrient enrichment that is dominated by another pigment type (as judged by color of the culture). Once supplied with higher nutrient concentrations (i.e., when we switched to SN media), these PC-only strains were able to quickly out compete PE-containing strains that appeared to dominate originally. This highlights important questions about the factors that either allow clades to coexist or certain groups to dominate. In particular, little is known about the ability of certain strains to survive unfavorable conditions and how variation in this ability can impact clade distributions in nature.

4.4.3 Comparison of diversity from clone libraries and culture isolations

Some isolated clade representatives did not appear in environmental clone libraries and vice versa. Clades II, III, VI, VII, VIII, CB5, 5.2MVI, 5.2MV2, 5.2MV3, and 5.3I were isolated into culture, but did not show up in the clone library sequences. The reverse occurred for clade IV and subclades IA and IB, whose sequences were found in the clone libraries, but not in the isolated and sequenced strains. Ahlgren and Rocap [3] found a similar mismatch of diversity recovered from simultaneous culture isolations and construction of clone libraries for samples from the Sargasso Sea. This is perhaps not surprising given the potential biases in each method. Isolation procedures are likely to favor growth of certain clades over others, and we can surmise that subclades IA and IB and clade IV cells do not grow well or were outcompeted by other strains in our enrichment conditions. The strength of possible culturing biases is clearly demonstrated with the isolations of PC-only *Synechococcus* strains. These

were numerous in the culture collection, but almost absent in both the clone libraries and flow cytometry record. For the clone libraries, there may be primer biases that selectively favor the amplification of certain clades. The overwhelming dominance of clade I cells would also make rarer clades difficult to detect in the clone libraries. For each method though, rarefaction analysis suggests that our sampling yielded overall diversity estimates nearing a plateau. This leads us to conclude that further sampling with a given method is unlikely to provide much new information. It is also apparent that in combination, culture isolation and clone library techniques have captured a more complete picture of total diversity than either one alone (Fig. 4-13).

It is noteworthy that the occurrence of certain clades differed among enrichments that were separated only by a few weeks in time. For example, the samples for enrichments 15-17 came from similar nutrient and temperature conditions, but each of these yielded a very different array of clade representatives (Fig. 4-11). While there are many biases in culturing, these differences between enrichments may hint at fast changing dynamics either in the field or in the enrichment culture (probably during the first few weeks). This illustrates some of the challenges faced when attempting to isolate and culture novel strains of *Synechococcus* or other organisms. We do not yet understand all the factors that determine how an organism will grow in isolation under laboratory conditions, and caution is needed in extending findings to better understand natural dynamics.

This research demonstrates that the *Synechococcus* population on the New England Shelf is mainly composed of cells that contain PE as their light harvesting pigment. PC-only *Synechococcus* appear to be rare and not endemic to this site, despite a high isolation frequency. In terms of diversity, clade I is the dominant clade throughout the year at MVCO, but with notable variability in the relative abundance of distinct subclades over an annual cycle. The *Synechococcus* population at MVCO is diverse, with 13 different clades observed. Many of these clades are relatively rare and appear only during late summer or early fall, suggesting that conditions are favorable for them only at this time of year. We emphasize the importance of investigating the temporal aspects of diversity patterns in temperate coastal systems. Many questions remain not only concerning how environmental (temperature, light, nutrients) and ecological (grazers, viruses, heterotrophic bacteria) factors govern clade distributions, but also how changes in these factors over a seasonal cycle affects the abundance of different clades. On the New England Shelf, *Synechococcus* cell concentration undergoes a dramatic (3 orders of magnitude) seasonal cycle, and it is likely that some of the abundance patterns are determined by which clades are favored

under different conditions. With isolates in culture from this location, we are poised to begin exploration into some of the differences between clade and subclade types. Ultimately, high frequency monitoring of clade diversity, coupled with physiological and ecological knowledge of representative strains, will allow a greater insight into how diversity of this genus is maintained and how that diversity is linked to overall population dynamics.
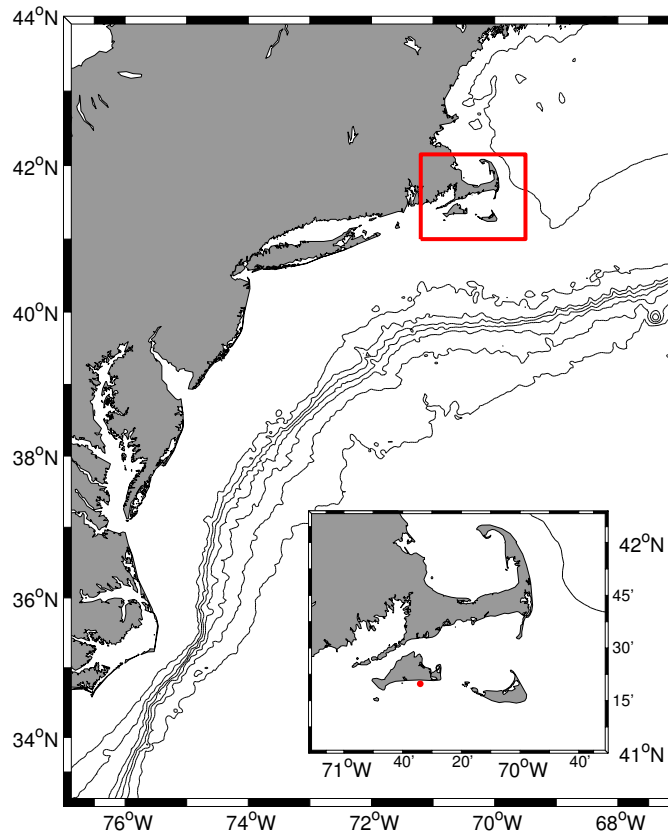
Figure 4-1: Map of Martha's Vineyard Coastal Observatory (indicated by red dot on inset map) and surrounding shelf area. Bathymetric contours are shown for 100 m, and for 500 m to 3000 m in 500 m intervals.
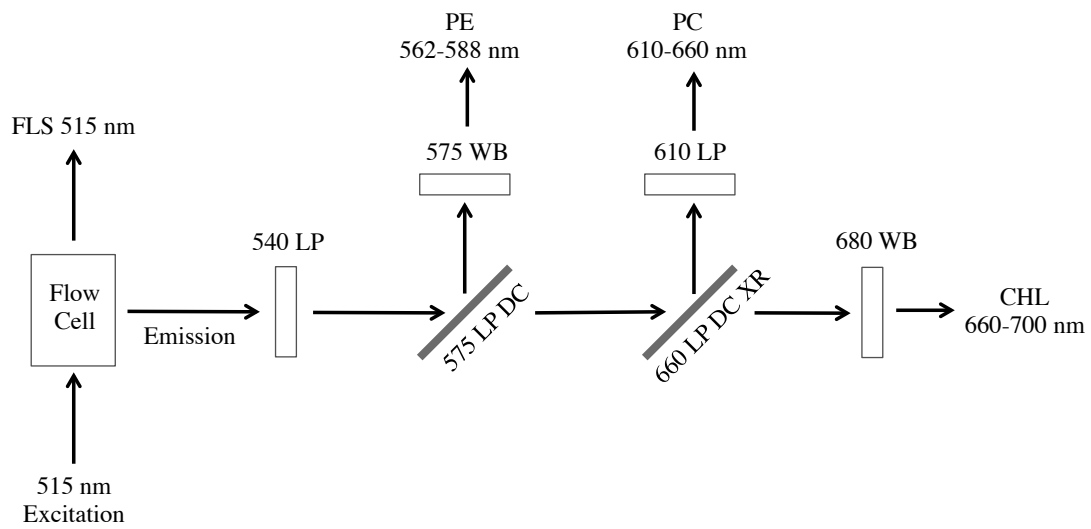
Figure 4-2: Schematic of optical configuration used for fluorescence detection in the EPICS V flow cytometer. Long-pass filters (LP) transmit light of longer wavelengths; long pass dichroic (DC) mirrors transmit longer wavelengths and reflect shorter ones. Wide band (WB) filters transmit light in a wavelength band centered on the designated wavelength.
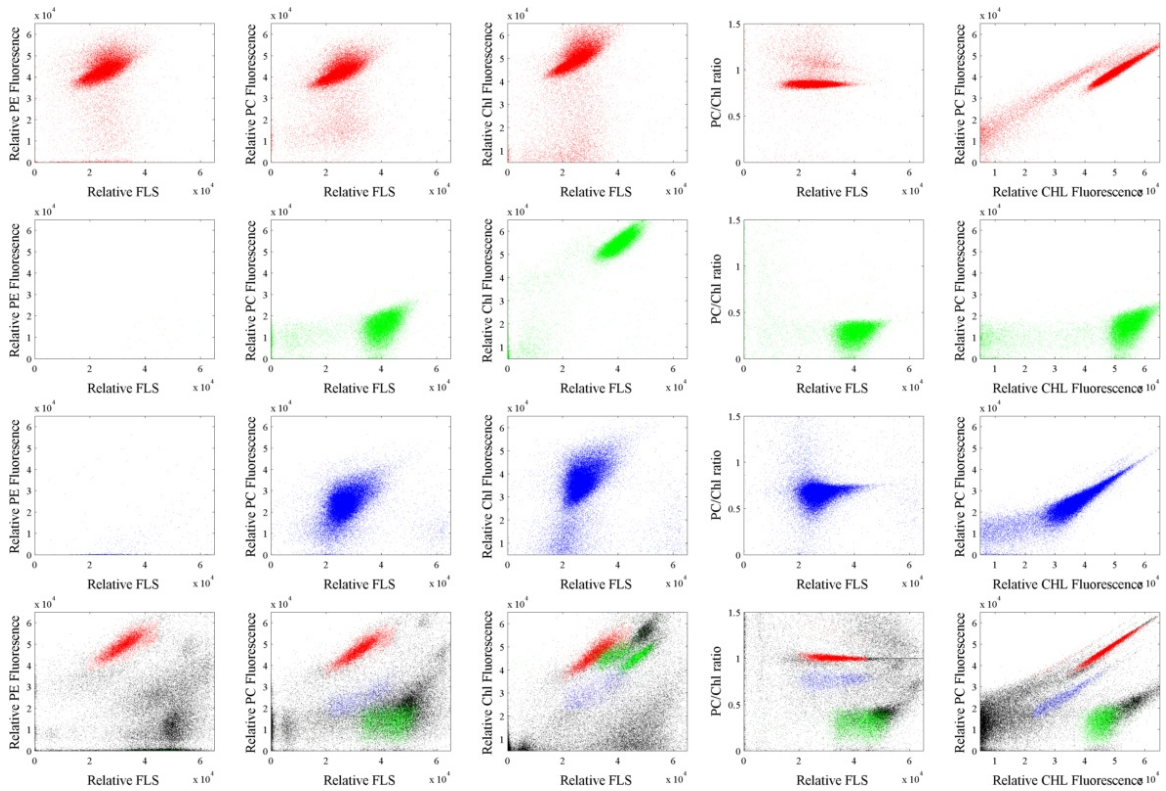
Figure 4-3: Cytograms of a PE-containing *Synechococcus* (strain MV1001, first row), a picoeukaryote (*Micromonas* sp.; 2nd row), PC-only *Synechococcus* (strain MV1312, 3rd row) and a field sample (October 11, 2010, 4th row). Columns show combinations of relative fluorescence and forward light scattering (FLS) as follows: (1) PE vs. FLS; (2) PC vs. FLS; (3) Chlorophyll (Chl) vs. FLS; (4) PC:Chl vs. FLS; (5) PC vs. Chl. For each sample, PE-containing *Synechococcus* are denoted by red, picoeukaryotes by green and PC-*Synechoccoccus* by blue. Unidentified larger phytoplankton or other particles are in black for the field sample.
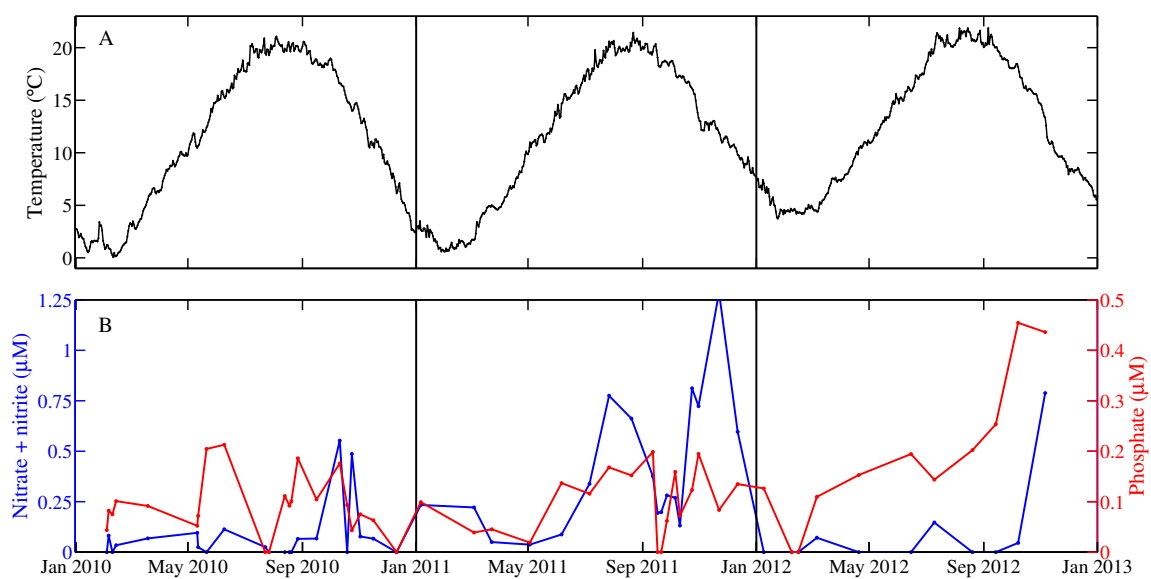
Figure 4-4: A. Water temperature (°C) at 4 m depth for 2010-2012. B. Concentration of nitrate + nitrite (μM) (blue) and inorganic phosphate (μM) (red).
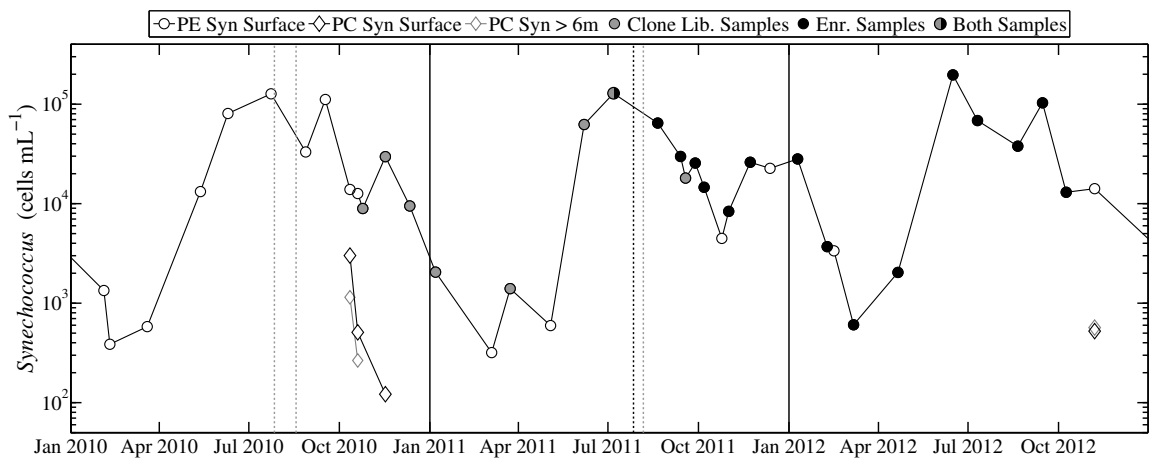
Figure 4-5: Cell concentration of PE-containing *Synechococcus* and PC-only *Synechococcus* for 2010-2012 at MVCO. PE-containing *Synechococcus* are represented by open circles, and PC-*Synechococcus* by diamonds. Color of circle markers indicate samples used for either clone library construction (grey), culture enrichments (black), or both (half and half). Dotted vertical lines indicate dates when samples were taken for either clone libraries or culture enrichments (depending on respective color), but not for flow cytometric analysis. Solid black lines indicate start of years.
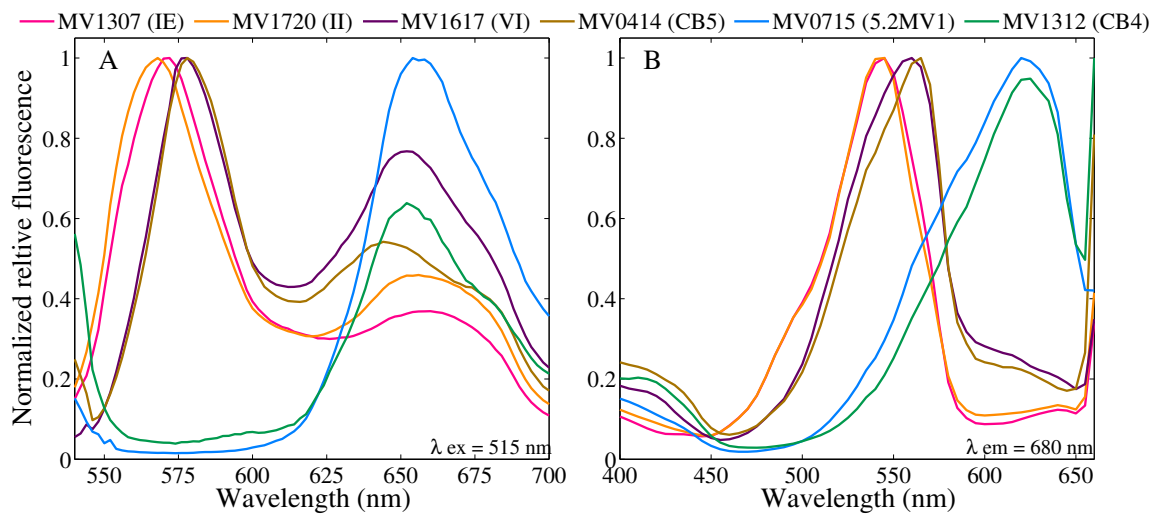
Figure 4-6: A. Selected normalized emission spectra for 6 different isolates (excitation at 515 nm) indicating presence or absence of PE (characteristic peak 550-580 nm). B. Normalized excitation spectra for chlorophyll emission at 680 nm for the same isolates, confirming presence or absence of PE. Strains MV0715 (blue curves) and MV1312 (green curves) were shown not to contain PE, whereas the spectra for other strains showed characteristic PE emission and excitation peaks. The non-PE containing strains exhibited spectra dominated by PC excitation and emission in the range 600-650 nm. Strains belong to the following clades: IE for MV1307, II for MV1720, VI for MV1617, CB4 for MV1312, CB5 for MV0414, and 5.2MV1 for MV0715.

Figure 4-7: Cultures of representative strains for each clade and subclade isolated from MVCO. Two strains are shown for each of clades II and 5.2MV1 to demonstrate range of color differences. See Figure 4-6 for representative excitation and emission spectra of clades IE, II, VI, CB4, CB5, and 5.2MV1.
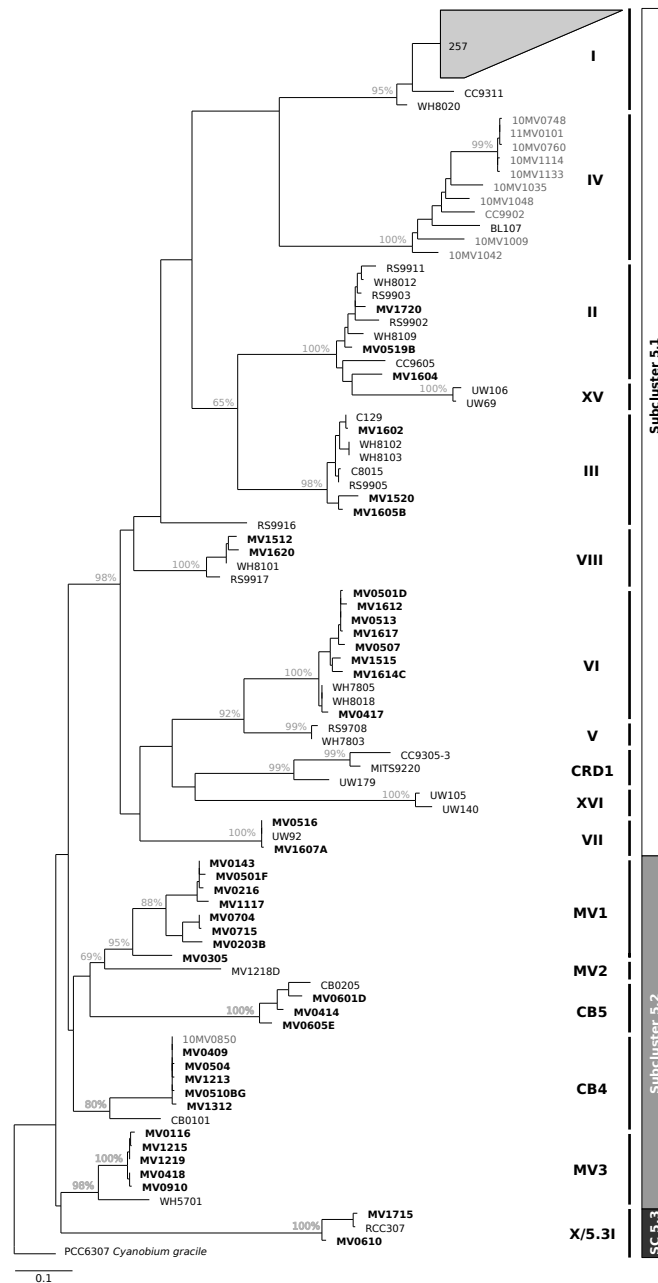
Figure 4-8: Phylogenetic tree constructed from *ntcA* sequences, illustrating the relationships between known clade representatives (in black), clone library sequences (gray), and culture isolates (bold font). Clade assignments for sequences were made by identifying the closest known clade representative or were assigned to possibly novel clades (designated as 5.2MV1, 5.2MV2, and 5.2MV3) if sequences did not cluster with known strains. Bootstrap values greater than 65% are shown on branches.

Figure 4-9: Expanded clade 1 branch of the *ntcA* phylogenetic tree in Fig. 4-8. Different subclades (OTUs from distance metrics) are labeled. Subclade IA is in dark blue, subclade IB is highlighted in medium blue, subclade IC is in black, and subclade IE is in cyan. The two clade I reference strains (WH8020 and CC9311) are highlighted in grey and form subclade ID. Bold font indicates a culture isolate from MVCO. The branch is split into three consecutive sections for ease of viewing.

Figure 4-10: Relative abundance of *ntcA* clone library sequences belonging to 7 distinct OTUs for 11 samples taken between July 2010 and September 2011. Shades of blue indicate OTUs that belong to clade I, orange/yellow indicate OTUs that belong to clade IV, and green indicates the OTU that belongs to clade CB4. The number of sequences retrieved for each library are indicated on the top axis.

Figure 4-11: Occurrence of each color type in *Synechococcus* cultures from the 17 different enrichments. Each row represents either a clonal isolate or an enrichment tube that was not selected for further isolation steps. For sequenced isolates, the clade designation is noted. Enrichments or isolates not sequenced are denoted by NS ('not sequenced'). Colors are approximate pigment color representations (see Fig. 4-7).

Figure 4-12: Nitrate + nitrite concentration (μM) and temperature ($^\circ$C) at the time of sampling for isolates belonging to 12 different clades. Each panel shows results for one clade. Each symbol represents an enrichment from which a clade representative was isolated and the color indicates the number of isolates per clade per enrichment. The bottom right panel shows the combinations of nitrate + nitrite and temperature that were encountered during sampling.

114
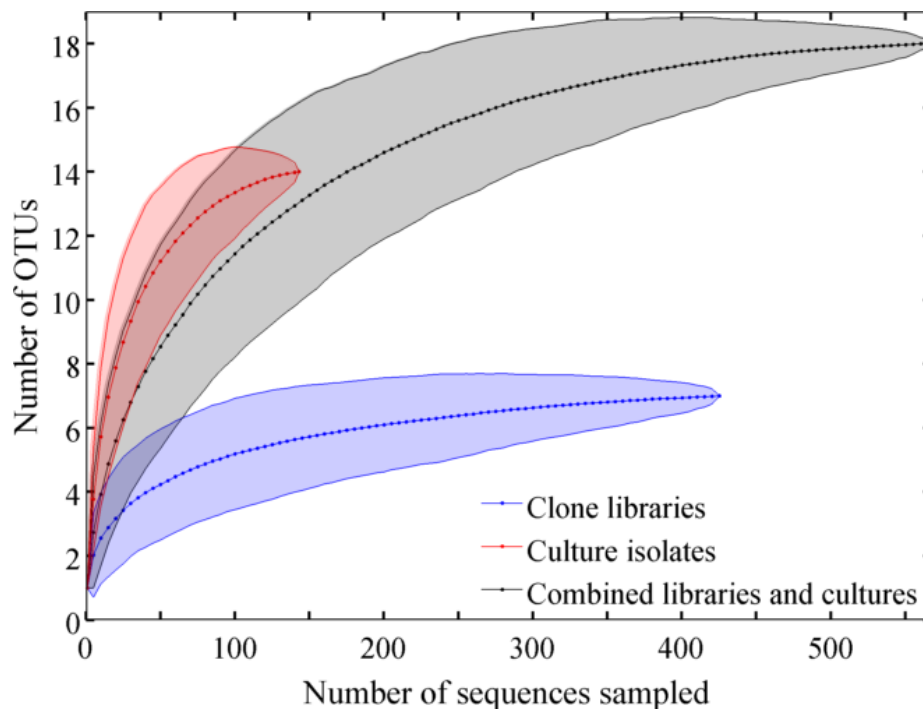
Figure 4-13: Rarefaction analysis of *ntcA* clone library sequences and culture isolate sequences, with OTU designation at 10% dissimilarity level. Dotted lines are the average number of OTUs observed for 1000 sampling iterations for a given number of sequences sampled. Shaded areas indicate the 95% confidence intervals.

Chapter 5

Seasonal *Synechococcus* diversity patterns and their relationship to population dynamics

5.1 Introduction

Changes in cell abundance for any phytoplankton population are the combined result of cell division, cell losses, and advection and mixing over time. A major challenge is to understand the contribution of each process to changes in cell concentration. A further layer of complexity, however, lies in the interaction between these elements and the physiological diversity present within the population. The dynamics at any one time are governed by the underlying physiological and ecological properties of the cells in the population. Properties such as temperature growth response, light harvesting capability, nutrient acquisition, and many others will affect the rate of cell division. Ecological differences such as palatability to heterorophic grazers or resistance to viral infection will affect cell loss rate. Diversity in these attributes can have a tremendous impact on the ability of cells to survive and thrive in a given environment. In temperate waters, where environmental properties undergo large seasonal changes, how diverse a population is may be particularly important for the population as a whole to be successful at any given time.

This appears to be the case for *Synechococcus*, a genus that has a large amount of genetic diversity that is thought to be representative of physiological or ecological diversity [92]. This genus has been partitioned into 20 well-defined clades based on molecular phylogeny [24, 92]. Physiological and biogeographical studies suggest there

are differences among the clades in terms of preferred environments and physiological capabilities [80, 68, 3, 120]. Questions remain, however, as to how this observed genetic diversity translates into physiological and ecological characteristics, where ultimately interactions with and selection for diversity takes place. Important to this understanding are studies that investigate the diversity patterns over time. Such studies have demonstrated shifting dominance of different clades, as well as clade successions over a seasonal cycle. In California coastal waters, Tai and Palenik (2009) found that clades I and IV were always dominant, but with changing relative abundance over the seasonal cycle. In the Gulf of Aqaba, Post et al. (2011) observed a succession of clades across the transition from winter mixing to summer stratification, which led to insights of possible preferred nutrient environments for clades I, III, and V/XII. Studies such as these allow insight into the environmental factors that drive the *Synechococcus* population over time.

The Martha's Vineyard Coastal Observatory (MVCO) offers an excellent venue to explore the relationship between population diversity and population abundance dynamics. A custom built, automated flow cytometer (FlowCytobot; FCB, [74]) allows high resolution (hourly) data on cell abundance for extended periods of time. From this instrument, we also obtain a time series of cell size distributions. These, together with a matrix population model, can be used to estimate an in situ division rate of the population as a whole (see Chapter 2). At MVCO, cell abundance and division rate follow a repeatable seasonal pattern. Abundance is low during winter (few hundred cells mL^{-1}), but increases rapidly during the spring in an annual bloom. This three-of-magnitude increase takes place in the span of a few months. The spring bloom results from increased division rates (coincident with increasing temperature), but how and when abundance increases appear to be a function of loss processes. Cell concentration typically stops increasing at around 10^5 cells mL^{-1} at the beginning of June. Division rates and cell abundance remain high for the summer. Division rate begins to decrease quickly in late fall, possibly from combined temperature and light limitation. Lower division rate coupled with still high losses results in a decline of cell abundance for the late fall and winter. Overlaid on this general pattern of cell abundance are also large cell concentration increases and decreases that occur with a period of approximately 2-4 weeks. These events are especially apparent in summer, when short term changes in cell concentration can be up to an order of magnitude.

We do not yet know how these cell abundance and division rate patterns relate to the underlying diversity structure. Investigations of the *Synechooccus* population at MVCO (see Chapter 4) so far have revealed that the population is, indeed, diverse.

Despite a large number of different clades found, the population appears to be dominated by clade I, with the possibility of subclades of this clade being important at different times of the year. Some clades have been isolated only during late summer and fall suggesting they may only grow during this time. To further explore these tentative patterns more quantitatively, we undertook sequencing of the V6 region of the 16S rRNA gene of the entire bacterial community via the Illumina sequencing platform. This method provides quantitative relative abundance estimates of *Synechococcus*. While the V6 sequences are very similar within a genus, small nucleotide differences can resolve tag sequences into different clades or groups of clades [84]. This approach, combined with the data and knowledge of the population dynamics, allows us to investigate not only how the diversity structure may change over the annual cycle, but also how this structure may affect the overall population dynamics.

We find that the population structure is composed of five main *oligotypes*, which are present year-round, but differ in relative abundance over the annual cycle. Certain oligotypes dominate the spring bloom, while other types become important later in the summer and fall. This data, coupled with information about cell abundance and division rate, suggest that the underlying diversity structure is critical in regulating the observed population dynamics.

5.2 Materials and Methods

5.2.1 Sample collection and DNA extraction

Seawater samples were collected near the MVCO offshore tower (41°19.500' N, 70°34.0' W) at roughly bimonthly to monthly intervals over a 3-year period from August 2010 to October 2013 for a total of 53 samples. Water was sampled at the surface via bucket sample or at 2 m depth with Niskin bottles attached to a rosette sampler on board the R/V Tioga. Two-three L of surface seawater was prefiltered through a 20 μm Nitex® mesh and then filtered onto 0.2 μm Sterivex® cartridge filters (Millipore) under vacuum pressure of no more than 40 kPa. Approximately 1.8 mL of DNA cell lysis buffer (Qiagen) was added to each cartridge before freezing and storage at -80 °C. For DNA extraction, samples were thawed on ice and approximately 200 μL of 0.5 mm zirconia-silica beads (BioSpec Products) were added to the cartridges. These were shaken vigorously at 2500 rpm for 10 min to break open cells. Continued DNA extraction followed a modified procedure with Qiagen Purgene reagents as described in [78]. DNA concentration was determined with a NanoDrop 2000 spectrophotome-

ter (ThermoScientific).

All sequencing was preformed at the Keck Facility within the Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA.

5.2.2 Bacteria community profiling via sequencing of the V6 hypervariable region

The V6 region of the 16S rRNA gene (ca. 60 bp) contains enough variance to be able to identify the bacterial genus [39]. This region was amplified for each environmental MVCO sample in two successive PCR reactions. Reactions were performed in triplicate, each containing 5-20 ng of DNA template, 0.2 μM unfused V6 amplification primers (967F and 1046R, [86]), 1 unit of Platinum$^{®}$ Taq DNA Polymerase High Fidelity (Invitrogen), 2 mM MgSO$_4$ (Invitrogen), and 0.2 mM dNTPs (Bioworld) in a total 33 μL volume. Reactions were performed on a GeneAmp PCR System 9700 thermocycler (Applied Biosystems) with the following conditions: 94 °C for 3 min; followed by 25 cycles of 30 sec at 94 °C, 45 sec at 60 °C, and 1 min at 72 °C; with a final extension step of 72 °C for 2 min. Presence of positive products (an expected 106 bp length) was checked on a electrophoresis microfluidics platform (LabChip GX, Caliper LifeSciences). Products were pooled and cleaned with a Qiagen MinElute PCR purification kit and rechecked on a LabChip GX after cleaning to ensure product was intact. Cleaned product served as the template for the second PCR reactions containing primers fused to unique barcodes compatible with the Illumnia HiSeq platform (bridge adapter, sequencing primer binding sites). Five μL of product was used as template for identical reactions as described above, but with unique barcoded primers for each sample. For each unique reaction (i.e., different barcoded primer), a no-template control reaction was also run. Cycling conditions were the same, with the exception that only 5 cycles were run with a final extension of 10 min at 72 °C. Confirmation of products and approximate concentration of each product was determined again on a LabChip GX. Products were pooled in equimolar amounts and cleaned with a Qiagen MinElute PCR purification kit. Products were size selected on a Pippin Prep (SageScience) and quantified with qPCR (Kapa BioSystems) to measure final concentration before sequencing on 30% of a lane on an Illumina HiSeq. Reads were demultiplexed based on the combination of index (CASAVA 1.8) and barcode with custom python scripts from the Bay Paul Center. Taxonomy was assigned to reads according to the Global Assignment of Sequence Taxonomy (GAST) with RefSSU, a primary reference database of near full-length

reference sequences, (derived from the SILVA rRNA database project (version 95)) within the VAMPs pipeline [40].

5.2.3 Sequencing of 16S rRNA gene from MVCO culture isolates

To investigate how the identified *Synechococcus* V6 tags relate to strains isolated from MVCO, the 16S rRNA genes of selected *Synechococcus* isolates from MVCO were sequenced to obtain the V6 region. Thirty-two cultures were chosen to represent the diversity of clades found at MVCO (see Chapter 4, Table A.2). Approximately 2 mL of dense culture isolate was centrifuged at 10,000 rpm for 6 minutes to pellet cells and DNA was extracted from this pellet with a DNeasy Plant Kit (Qiagen), following manufacturer's instructions with the exception of final elution volume (75 μL). Approximately 20 ng of DNA was added to PCR reactions with 107F and 1313R cyanobacteria-specific 16S rRNA primers (Fuller et al. [28]). Final primer concentration was 2.5 μM in a total reaction volume of 50 μL with Qiagen Taq PCR Master Mix Kit reagents. BSA was added at 0.2 mg mL^{-1} final concentration. Reactions were preformed on a GeneAmp PCR System 9700 thermocycler (Applied Biosystems), with an initial denaturation period of 4 min at 94 °C; followed by 30 cycles of 1 min at 94 °C, 1 min at 55°C, 1 min at 72 °C; and then a final extension step at 72 °C for 7 min.

Product was cleaned with a MinElute PCR Purificaiton kit (Qiagen). Cleaned products were cloned into TOPO vectors for sequencing (TOPO TA Kit, Invitrogen) and transformed into chemically competent *E. coli* TOP10 cells (Invitrogen) following manufacturer's instructions. At least 5 positive colonies (determined by blue/white selection on X-Gal, kanamycin LB plates) were picked. Plasmids were obtained via automated plasmid purification with a BiomekFX at the Josephine Bay Paul Center Keck Facility. Due to the length of the 16S rRNA gene (> 1000 bp), two sequencing reactions were set up to sequence the insert from both ends of the vector. Sequence reactions used BigDye Terminator chemistry (Applied Biosystems) and contained at least 200 ng plasmid. One reaction contained M13 reverse primer, while the other contained M13 forward primer (15 μM, both from TOPO kit). Sequencing was performed on 3730 DNA Analyzer (Applied Biosystems). Primer and vector sequences were removed and resulting sequences were identified by BLAST search against 16S rRNA accessions in NCBI Genbank. For each sample, the forward and reverse sequences with positive identification to *Synechococcus* were joined to obtain the full

length 16S rRNA gene and aligned with ClustalW in BioEdit (version 7.2.0, [33]). V6 region was identified by alignment with other V6 sequences.

5.2.4 Clade identification of *Synechococcus* V6 environmental tags from reference strains

We constructed a comparison database of *Synechococcus* reference V6 sequences by searching the NCBI Genbank database for 16S rRNA gene sequences from established *Synechococcus* strains. Only full length sequences that contained the V6 region and for which unambiguous clade designation could be made were included. Strain clade designation had usually been determined from higher resolution diversity markers (i.e., 16S ITS, *ntcA*, *petB*). A total of 156 sequences were found and the V6 portion of these sequences, as well as those of the MVCO culture sequences, was used to relate and categorize the V6 environmental tag sequences to known clades or groups of clades. An environmental tag was checked against this database for a match if it was found more than 200 times in the environmental sequences. Some tag sequences could not be matched to known strain sequences, and for these cases, the closest similar sequence was used to infer a possible identity. Distance between tags was computed with mothur v.1.23.1 [94] as uncorrected pairwise distances.

5.2.5 *Synechococcus* cell concentration from FlowCytobot

FCB was deployed at the MVCO offshore tower, sampling 4 m below mean water level. Details of the design and performance of FCB are described elsewhere [74]. Data analysis and enumeration of *Synechococcus* cells were as described in [98].

5.2.6 Population division rates from cell size distributions

We determined daily division rate at MVCO from application of a matrix population model to the 3-year time series of cell size distributions obtained from FCB (FCB measures side angle scattering which is converted to cell volume [74]). Details of model and model application are in Chapters 2 and 3.

5.3 Results

5.3.1 Resolution of *Synechococcus* clades from V6 sequences

From the 156 reference V6 sequences, along with the 32 sequences from the MVCO cultured isolates, we identified only 19 unique V6 sequences. Differences in sequences only occurred at 17 nucleotide positions, and the maximum difference between any sequence was 10 nucleotides, but often sequences differed by 3 - 5 nucleotides. For some clades, reference strains did not share the same V6 sequence, and certain strains had identical sequence matches to other strains belonging to different clades (i.e., more phylogenetically distant) (see Fig. 5-1). V6 sequence differences within a clade were usually at one nucleotide position. Interestingly, clade I tags appeared to be very distinct, with a 'TT' at position 34 and 35 in the V6 region, so tag and clade identity can be unambiguously linked for this group.

5.3.2 *Synechococcus* environmental tags and clade identification

Sequencing of the V6 hypervariable region via the Illumina platform yielded a total of 17,802,457 tags for the entire bacterial community from 53 surface water samples. Of these tags, 258,984 were classified as *Synechococcus*. In general, the number of *Synechococcus* reads normalized to total reads tracked the *Synechococcus* cell concentration well (Fig. 5-2A). A total of 15 unique environmental tags were found to exceed our threshold of at least 200 sequences in the dataset. Of these 15 tags, 10 had positive matches to at least one or more known strains. The top 5 accounted for 89% of the total *Synechococcus* reads. When plotted on a log-log scale, the number of *Synechococcus* reads normalized by total reads per sample correlated well with total *Synechococcus* abundance for cell concentrations above 10^4 cells mL^{-1} (see Fig. 5-2B).

The first and fourth most abundant tags matched only to clade I sequences (labeled here as O1-I and O4-I to designate tag rank and clade match). Tags O1-I and O4-I each map to distinct clusters within the clade I branch of the 16S rRNA tree (Fig. 5-2). When we consider the phylogenetic relationships of isolates within the *ntcA* tree (see Chapter 4), this correspondence to divisions within clade I was not 100%. Most of the O1-I and O4-I V6 tags correspond to subclades IC and IE, respectively, but there were a few exceptions that complicate interpretation of how these tags map to the different subclades of clade I. Tag O1-I matched to MVCO strains belonging to subclade IC, but also to two MVCO strains belonging to subclade IE (strains MV1112 and MV0908).

Only two reference strains matched this tag sequence, Almo3 and SYN20. Because we do not have *ntcA* sequences for theses strains, we cannot assign these strains to a sublcade with certainty, but it is likely that these strains are clade IC, as they cluster with the MVCO IC strains in the 16S rRNA tree (Fig. 5-1). Tag O4-I matches strains belonging to subclade IE (both MVCO strains and reference strain WH8016), but also strains belonging to other subclades (i.e., subclade ID: WH8020 and CC9311, see Chapter 4). The exact details of tag to subclade correspondence cannot be fully resolved, but taken together this information leads us to believe that these two clade I tags indicate distinct genetic types within the population at MVCO.

Though no exact matched was detected, the fifth most abundant tag also likely belongs to clade I (labeled O5-I*). In cases with no exact match, the reference tag with greatest similarity was used for a likely clade designation (see Table 5.1). An additional clade I tag is consistent with the presence of one other subclade at MVCO (subclade IA). Clone libraries of *ntcA* revealed the presence of a total of 4 different subclades, only two of which are represented in our culture isolates. Other unidentified tags, lower in abundance, were also observed that likely belong to clade I.

The second and third most abundant tags matched to multiple clades (labeled as O2-M and O3-M, for 'multi'). For these two tags, clade designation was not possible. The O2-M tag had identical matches to strains belonging to clade II, III, IV, CRDI, VII, CB5, and others in the reference database and also matched MVCO strains belonging to clades III, VII, and CB5 (see Table 5.1). The O3-M tag matched strains belonging to clades II and XV and an MVCO strain belonging to CB5.

For tags with abundance ranked 6-15, some had unique matches to only one clade (VI, IX, 5.3I, CB4 and 5.2MV1; see Table 5.1), while others matched to multiple clades (II/VI, 5.3/III and VIII/XVI).

5.3.3   Cell abundance and division rate seasonal patterns

For a complete analysis of the cell abundance dynamics for years 2010 -2013 the reader is referred to Chapter 3. Briefly, cell abundance followed a repeatable seasonal pattern. Lowest cell concentration occurred in winter at a few hundred cells mL^{-1}. Concentration rapidly increased during a spring bloom to reach numbers that exceeded $2 \cdot 10^5$ cells per mL during summer. Cell numbers then declined slowly during late fall and winter back to low values. Overlaid on this seasonal pattern were variations that occurred on a shorter time scale (weeks-months). After the peak of the spring bloom, when cell concentration reached a maximum, typically concentration

then dropped to less than half of this value (from $2\text{-}3 \cdot 10^5$ to $1 \cdot 10^5$ cells mL^{-1}) (Fig. 5-3A). For the summer, cell concentration usually remained at or below 10^5 cells mL^{-1} and did not return to spring bloom peak values. Large changes in cell abundance occurred intermittently during summer into fall, often with events separated by 2-3 weeks.

Division rate also followed a repeatable seasonal pattern. Rates were low in winter and increased almost linearly during the spring. Rates remained relatively high during summer, but slowly decreased during the late summer and fall before returning to winter low values. Shorter timescale variation in division rate also occurred, particularly in the summer and early fall months. These can consist of multi-day declines followed by a return to moderate to high values.

5.3.4   Relative abundance patterns

Each of the five most abundant tags exhibited a repeatable seasonal pattern of relative abundance (Fig. 5-3D). Tag O1-I made up the majority of the *Synechococcus* population from late fall through the start of the next summer, usually peaking at the start of summer (40-50%), followed directly by a seasonal low of around 15-20% of the abundance. In late summer and fall, tag O2-M was usually dominant, and relative abundance reached a peak of 30-40% during this time.

For each year, tag O3-M reached a maximum relative contribution (20-30%) during the decline of tag O1-I, but before the maximum of tag O2-M. Tag O3-M did not change much in relative abundance throughout the year, and did not drop below 10% of the abundance at any time. The pattern of relative abundance of the other clade I tag, O4-I, was similar to that of tag O2-M. Relative cell abundance peaked in late summer and fall, while the minimum occurred in spring. The possible clade I tag, O5-I*, followed a similar pattern to that of tag O1-I; tag relative abundance peaked (14-20%) at the start of summer, and directly after reached its annual minimum ($\sim$3%).

All other tags, individually, made up less than $\sim 3\%$ of the population at any given time. Many of these tags did not show a seasonal pattern. An exception to this was tag O8-VI, which showed a repeatable peak in relative abundance in early fall (up to $\sim$1.5 %, see Fig. 5-4), which is consistent with culture isolations of this clade during this time of year (see Chapter 4 for more details). Tag O6-IX also appeared to peak in fall.

Environmental tags that matched to sequences of strains that do not contain

phycoerythrin (PE) (phycocyanin (PC)-only) were always low in abundance. The only tag in the top 15 that matched a PC-only strain was that of O10-5.2. This tag had very low relative abundance, with the exception of an episodic increase to ∼3% during November 2012. Tags that matched other PC-only *Synechococcus* strains isolated from MVCO were also low: totaling not more than 200 reads in the whole dataset. This low number of tags that match to PC-only *Synechococcus* is consistent with the flow cytometry record. Numbers of these cells are typically very low, with the exception of a few sporadic events (see Chapter 4). In general, the low number of tags supports this, with the episodic increase in tag O10-5.2 matching the timing of the increase observed in the flow cytometry record during November 2012. A similar match between the flow cytometer record and tag increase occurred for the tag matching MVCO strains from clade CB4. Here, tag relative abundance increased at the same time that PC-only *Synechococcus* cell number increased during October 2010 (see Fig. 5-4).

5.4 Discussion

5.4.1 Clade identification from V6 sequences

From analysis of the reference database and MVCO strain sequences, we find that the V6 region does not, in general, resolve distinct clades or even groupings of clades, (with a few exceptions; see Table 5.1). This level of ambiguity is perhaps not surprising since the length of the V6 region for *Synechococcus* is only 60 base pairs. This ambiguity does complicate interpretation of the environmental V6 *Synechococcus* tags that match to multiple clades. These tag sequences may be composed of just one clade or be a composite of several. A possible clade identity could be inferred from the direct matches to MVCO strains or clades observed in clone library data (see Chapter 4). Tag O2-M is then likely to be composed of clade II, IV, VII and/or CB5, and O3-M composed only of clade CB5 (see Table 5.1). This assumption is subject to the biases in culturing and clone library construction, but offers likely candidates for tag identification. These clades are also likely candidates as they have also been observed in waters offshore of MVCO and in other coastal locations. Clade IV was observed by Ahlgren and Rocap [4] 80 km south of MVCO and in general is found in cooler, coastal waters [120, 106]). Clade II was also observed by Ahlgren and Rocap [4], but further south (304 km south of MVCO). Clade CB5 has been observed in coastal waters, as well as in Artic areas [18, 37].

The inability to resolve clades is a concern if clades matching to the same tag have different abundance dynamics. For the most part, we observe consistent seasonal dynamics of the relative abundance of these tags, suggesting that diversity resolved at the tag-level is ecologically relevant for understanding *Synechococcus* dynamics at MVCO. Despite not having definitive clade assignments, tag relative abundances provide information about how diversity patterns change at MVCO over an annual cycle. We therefore now refer to tags as *oilgotypes* to highlight the possibility of tags being ecologically relevant units.

5.4.2 *Synechococcus* diversity patterns

Changes in relative abundance of an oligotype could be the result of an increase in that cell type or the result of decreases of another oligotype (and vice versa). It is useful for interpretation of oligotype relative abundance to consider the changes in actual cell concentration. For example, if an oligotype demonstrates increasing relative abundance during a time of increasing cell concentration, it is likely that the relative abundance increase is due to an actual increase in those cells. Relative abundances scaled to cell concentration at the time of sampling demonstrate these relationships (see Fig. 5-5). While these numbers will not provide an error-free estimate of cell concentration for each oligotype (due to possible biases in sequencing pipeline), such an exercise is helpful to understand how the relative changes may relate to the actual cell abundance changes. From this, we find that the five main oligotypes can be partitioned into two categories: those that contribute to the spring bloom and those that constitute the summer/fall assemblage.

Spring bloom

The large, three orders-of-magnitude increase in cell abundance that constitutes the spring bloom appears to be the result of growth of three oligotypes: O1-I, O3-M and O5-I*. Oligotype O1-I constitutes the majority of cells during the spring bloom (up to 50%) and O3-M and O5-I* each compromise ∼20% of the population (Fig. 5-3). Both O1-I and O5-I* reach their maximum relative concentrations during this time. After cell concentration has stopped increasing, it declines dramatically, usually in July. During this time, cell concentration can drop to half or less of the peak values. This timing coincides with a decrease in the relative abundance of oligotype O1-I and O5-I*. Both viruses or heterotrophic grazers could be contributing factors to the decline of O1-I, as both are capable of targeting specific cells. For type O5-I*, spring

appears to be the only time when it is able to increase to a significant proportion of the population ($< 5\%$ for the remainder of the year). During the decline of O1-I and O5-1*, oligotype O3-M reaches its relative abundance maximum. This suggests that this maximum is not a result of increasing cells, but rather due to the decline of the other two oligotypes.

Late summer and fall

In August and September, O2-M becomes the dominant oligotype. Oligotype O4-I also reaches a relative maximum during the late summer (20-30%) and follows a similar relative abundance pattern to that of O2-M. The increase in relative abundances of these types is concurrent with secondary increases in cell abundance during this time. While cell abundance is still high ($\sim 10^5$ cells mL^{-1}), it is characterized by large (order of magnitude) fluctuations. Many of these increases in cell concentration occur when O2-M and O4-I also reach local maxima (despite a very noisy summer period). Sustained high cell concentration in the summer months and early fall may then be a result of increase in these oligotypes.

Late fall and winter

Oligotype O1-I reemerges as the most dominant oligotype around November of each year. O1-I remains the most abundant for the rest of winter, while the other oligotypes, with the exception of O5-I*, hover at 20%. Presence of these oligotypes at non-zero relative abundances suggests that they are able to persist in the environment during the unfavorable winter (and perhaps spring) conditions until they bloom again. Cell number declines to low levels (10^2-10^3 cells mL^{-1}) in December and January; the actual concentration of cells for each oligotype would be much less. Clade I isolates that matched to types O1-I and O4-I were isolated from winter waters, indicating that for at least these oligotypes, the cells are indeed viable and appear to over winter successfully. The ability of summer types to persist in the environment during winter is in contrast to the hypothesis proposed in Chapter 4. I suggested that clades only found during the summer and fall (i.e., II, VII) might be advected to MVCO and be unable to grow at other times of year. Transport from other environments and the ability to overwinter may both be occurring, but further interpretation is hindered by unresolved clade identity of the oligotypes. Given that types O2-M and O2-3 could be composed of multiple clades, it is possible that some clades are advected to and survive for a short while at MVCO, while others are able to persist

during the winter. Regardless, the presence of these 5 oligotypes during winter raises interesting questions about the ability of *Synechococcus* to survive harsh conditions. It is unknown whether these cells can enter a true dormancy state or simply undergo slow growth during this time. This aspect of cell physiology is important to understand as tolerance to cold temperature may determine which clades can persist at MVCO.

5.4.3   Potential causes of diversity shifts

The shift in dominance between oligotype O1-I and O2-M is a distinct and repeatable seasonal feature. Temperature may be an important factor. The spring bloom is a result of division rate increasing with temperature during the first half of the year (see Chapters 2). Division rates begin to increase when the water temperature reaches 5-6 °C and increases almost linearly up to ~ 1 d^{-1} when water temperature reaches 16-17 °C. To constitute the spring bloom, the cells belonging to oligotypes O1-I, O3-M and O5-I* must have the ability to grow at these rates in what is still relatively cool water. A preference for cooler water for the oligoytypes belonging to clade I (O1-I and O5-I*) may also explain their low relative abundances during summer. After July, O1-I relative abundance is negatively correlated with temperature (Fig. 5-6). Oligotype O2-M demonstrates the opposite pattern during the summer, when relative abundance has a positive correlation with increasing temperature. A possible preference for warmer temperatures is consistent with the biogeography of a few of the clades that could compose oligotype O2-M, such as clade II and VII, which have been found in subtropical and tropical waters [120]. Interestingly, we found no correlation between oligotypes and nutrient concentrations (nitrate + nitrite, phosphate and ammonium), suggesting that nutrients may not be a dominant environmental factor determining oligotype relative abundance in this coastal system.

Selective loss processes could be responsible for some of the relative abundance changes. Both heterotrophic grazers and viruses have the ability to remove specific cells from the system. This might, for instance, be the cause of the large decline of oligotype O1-I after the spring bloom. Heterotrophic grazers have been found to preferentially ingest different strains of *Synechococcus* [121, 6], such that certain clades may be preferred over others as prey items. Viruses may also regulate the composition of *Synechococcus* as the diversity of cyanophage has been shown to co-vary with that of *Synechococcus* over an annual cycle [71]. At MVCO, we have shown that loss and division rate are closely coupled throughout the year (see Chapter 3). For

Synechococcus, most of the biological losses are thought to come from grazers rather than viral lysis [112, 8, 110]. Micrograzers are capable of rapid response to changes in prey concentration, as their own division rates can match or exceed that of their prey [52]. This is not to exclude viral infection as a selective agent. Concentrations of *Synechococcus* that would enable increase of lytic phage have been calculated to be $\sim 10^4$ cells mL^{-1} [63], and studies of viral infections of natural populations are consistent with this estimate [88]. Above this concentration, susceptible populations should experience significant losses from phage. Our results suggest that the four most abundant oligotypes are all represent at $> 10^4$ cells mL^{-1} during most of the summer and fall. Thus they could be susceptible to pressure from viruses. It is important to note, however, that Waterury and Valois (1993) suggest that naturally occurring *Synechococcus* are largely resistant to their co-occuring phages, and that only a small percentage of the population is sensitive to infection. Resistance may come with division rate trade-offs such that sensitive cells have a growth advantage. These trade-offs could be important in determining relative abundance over the annual cycle.

5.4.4   Clade I diversity

The relative abundance patterns of oligotypes within clade I lead us to conjecture that there are substantive physiological and/or ecological differences among them. Here, we find that O1-I and O5-1* bloom during the spring while O4-I peaks during the late summer and fall. While both O1-I and O5-I* reach a maximum relative abundance at the same time, O1-I is present in much greater numbers, raising questions as to why this type is so successful during this time. Despite the general findings that clade I prefers temperate waters, tag O4-I peaks at the time of year when water temperature is highest. These seasonal timing differences in relative abundance support the idea that subclades within a clade can have their own environmental niches. This conclusion is also consistent with findings in other coastal systems. Tai et al. [105] found that subclades within clades I and IV showed distinct depth distributions along a coastal to open ocean transect, suggesting these subclades have preferred environments. It seems clear that not all members of a clade may respond in the same way to the environment, and further exploration of the microdiversity is needed. Going forward, we will be able to explore possible physiological and ecological differences between types O1-I and O4-I as we have representatives in culture. We were not, however, able to isolate a strain belonging to type O5-I*, suggesting that it may have

unique environmental preferences (or different culture and isolation requirements). The presence of other low abundance tags, likely belonging to clade I (Table 5.4.6), suggest there may also be other types which are similarly challenging to isolate (i.e., O7-I*, O12-I*, O14-I*).

5.4.5   Subpopulation model dynamics

The matrix population model used to estimate *Synechococcus* division rate includes two subpopulations that can have different physiological parameters. The estimated division rate is a composite of the division rates for the each of the two subpopulations. While the model is able to estimate whole population division rate accurately (see Chapter 2), it is unknown whether the model can resolve the division rates of subpopulations correctly. If a population contains more than one type of *Synechococcus*, the subpopulation division rates are then composites themselves. At MVCO, we have found that the *Syenchococcus* assemblage always consists of more than two types. Consistent with this, the model does not usually identify two identical subpopulations (i.e., the model parameters for each subpopulation are different). Typically, the two subpopulations also differ in their cell volumes. These two volumes follow a similar pattern to one another over the annual cycle: volumes are larger in late fall, winter and spring, and are smaller in summer. This matches the actual cell volume pattern observed from FCB data (presented in Chapter 3), and suggests that all cells in the environment follow the same phenomenon (i.e., each cell type changes size over the seasons).

How the model parameters relate to the actual cell volume dynamics of the different oligotypes is difficult to infer. Not only is the data inherently noisy (see Fig. A-35), but we cannot be sure which of the oligotypes are being represented by the two subpopulations (which are distinguished by their cell volume). Nonetheless, we observe a distinct change in the proportion ascribed to each of the model subpopulations during the time at which dominance between tags O1-I and O2-M switches (Fig. 5-7). We find that the subpopuation with smaller cell size contributes more than half of the model population when tag O1-I is dominant, but less than half when O2-M is dominant (Fig. 5-7). While it is tempting to try to investigate the subpopulation division rates that the model produces, we find that these rates can sometimes be much higher (2-3 d^{-1}) than actual division rates observed for *Synechococcus* cultures or field samples. While the model is able to accurately estimate population level division rates (Chapter 2), this suggests that it may not realistically represent all the

131

subopulation details. In addition, it is unclear if a given subpopulation represents the same oligotypes throughout the entire year. The model subpopulation with smaller cell size could represent one combination of oligotypes during part of the year and a different combination at other times. Regardless of these interpretation difficulties, the detection by the model of a clear shift in the underlying population structure indicates that the cell volumes dynamics of the population are indeed different at this time of year and point to physiological differences between the oligotypes.

5.4.6   Conclusions

From the V6 sequence data, we find that the *Synechococcus* community is composed mainly of 5 oligotypes. While all are present year-round, they do differ in their relative abundances in a repeatable seasonal fashion. Two types switch dominance during the year; O1-I (likely representing subclade IC) dominates from late fall through spring, while O2-M (likely composed of clade II, VII and/or CB5) becomes dominant in summer and fall. There are two broad categories: oligotypes that bloom in spring and those that are abundant in late summer and fall. The spring bloom appears to be a direct result of three oligotypes that are able to increase in abundance during this time. The sustained summer and fall cell concentrations and division rates appear to be a result of the two other oligotypes. This pattern of relative abundance, coupled with our knowledge of cell concentrations and in situ division rate, imply that the overall *Synechococcus* population dynamics are strongly affected by the underlying diversity structure. This highlights the need to understand the environmental and ecological factors that determine which clades are able to survive and grow at MVCO. In this system, temperature likely plays an important role in determining when the different oligotypes are favored. Top down control either by grazers or viral lysis may also be important in determining the relative abundance patterns. Future research should attempt to identify clade or subclade composition of the oligotypes, as we have culture representatives of many clades from MVCO. Exploring possible physiological and ecological differences between the clades will be important for refining general clade niches and preferred environments. Ultimately, knowledge of the physiology and ecology of representative isolates, together with abundance data of each clade, will allow us to fully understand how the cell abundance changes are linked to and determined by the diversity present at any given time.
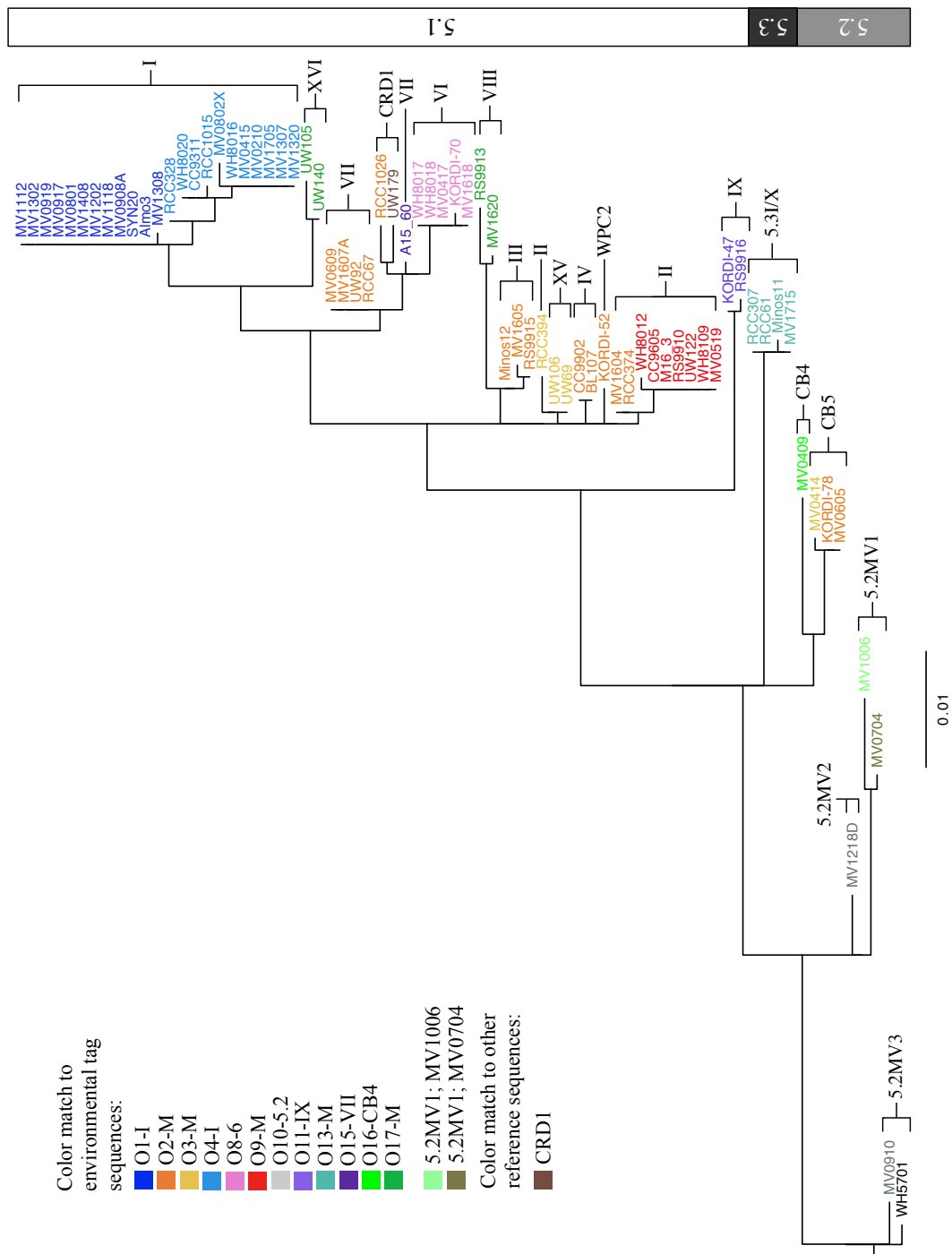
Figure 5-1: Phylogenetic relationships between selected clade representatives constructed with full length 16S rRNA sequences. Color of strains match those in Table 5.1 and indicate strains that have the same V6 sequence to environmental tags. White and grey vertical boxes indicate subcluster that clades belong to. Phylogenetic reconstruction was carried out in MEGA (version 5.2.2, [107]), with a maximum likelihood approach and a GTR GAMMA rate substitution model. Positions were not used if there was a gap in the aligned sequences for the majority of strains.

133
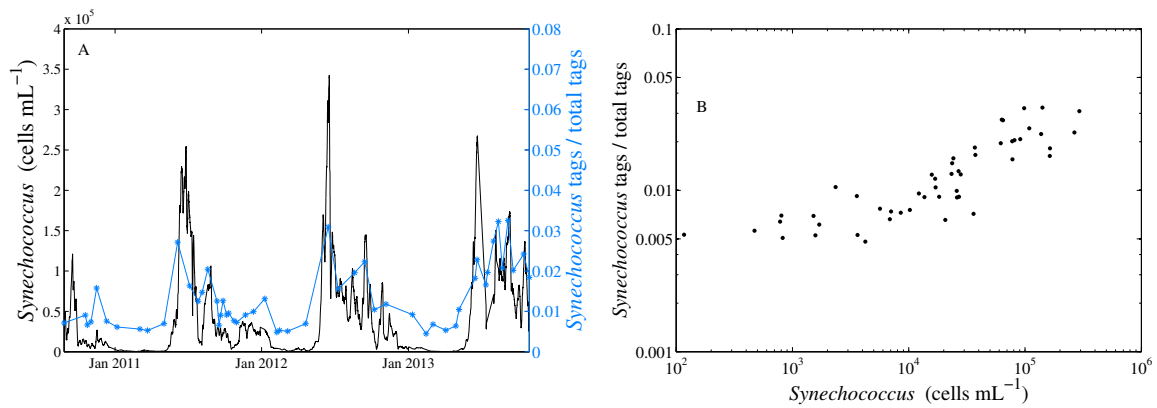
Figure 5-2: A. *Synechococous* cell concentration determined from FCB (black line) overlaid with normalized *Synechococcus* reads (light blue). B. Relationship between *Synechococous* cell concentration (log scale) and normalized *Synechococcus* reads (log scale). There appears to be a linear relationship between the log normalized reads and log cell concentration after 10^4 cells mL^{-1}.
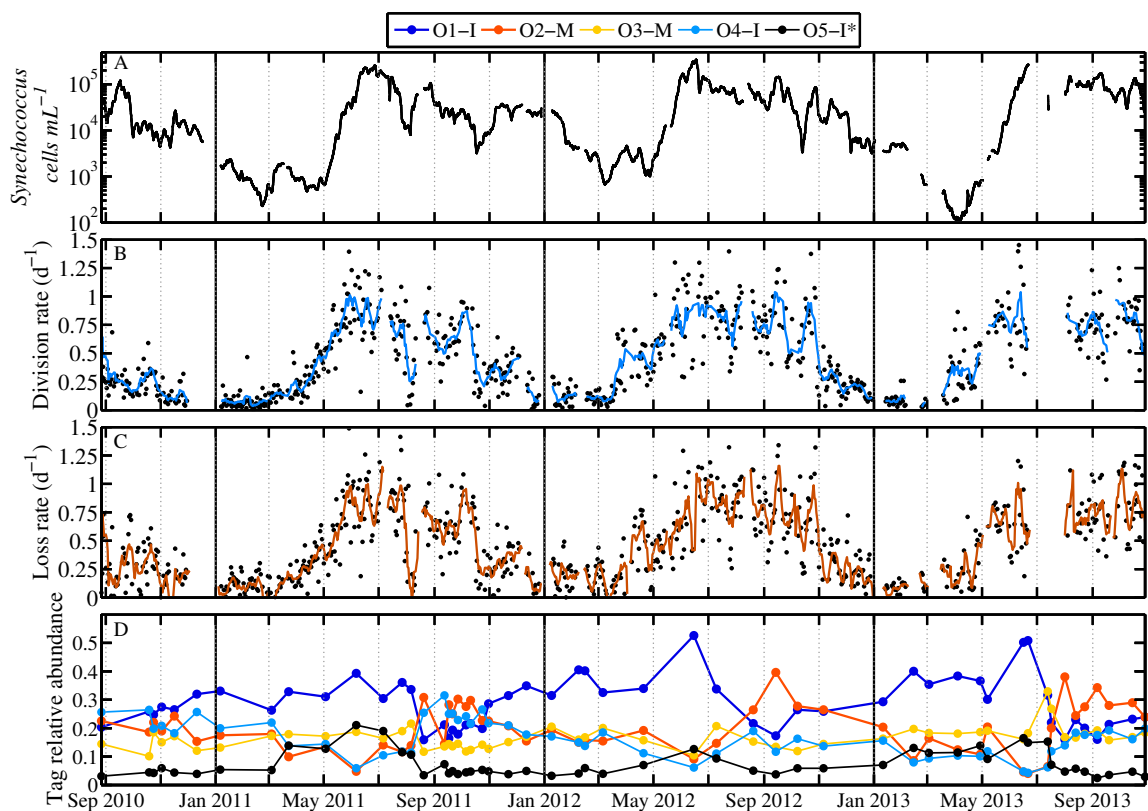
Figure 5-3: A. *Synechococcus* cell abundance from Aug 27, 2010 through Oct 29, 2013 (log scale). B. Division rate estimates during same time. Blue line denotes 7-day running average. C. Loss rates calculated from division rate and net growth rate (obtained from 48-hour smoothed cell abundance). Red line denotes 7-day running average. D. Relative abundance of V6 tags of the 5 most abundant tags. Clade I tags are denoted in shades of blue and other tags are denoted in orange and yellow. Solid black vertical lines indicate January 1 of each year.

135
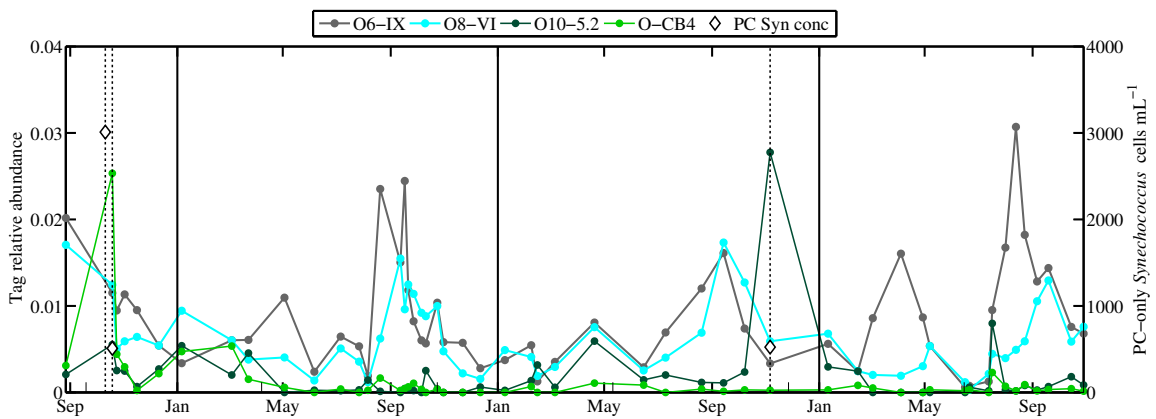
Figure 5-4: Relative abundance of tags matching to clades IX (gray line) and VI (cyan line) and to strains that contain PC as their light harvesting pigment: 010-5.2 (dark green line) and O-CB4 (light green). Diamonds indicate cell concentration of PC-only *Synechococcus* as determined in Chapter 4. Dotted lines are for convenience to indicate when PC-only *Synechococcus* cells were observed in the flow cytometry record. Solid black vertical lines denote start of each year.
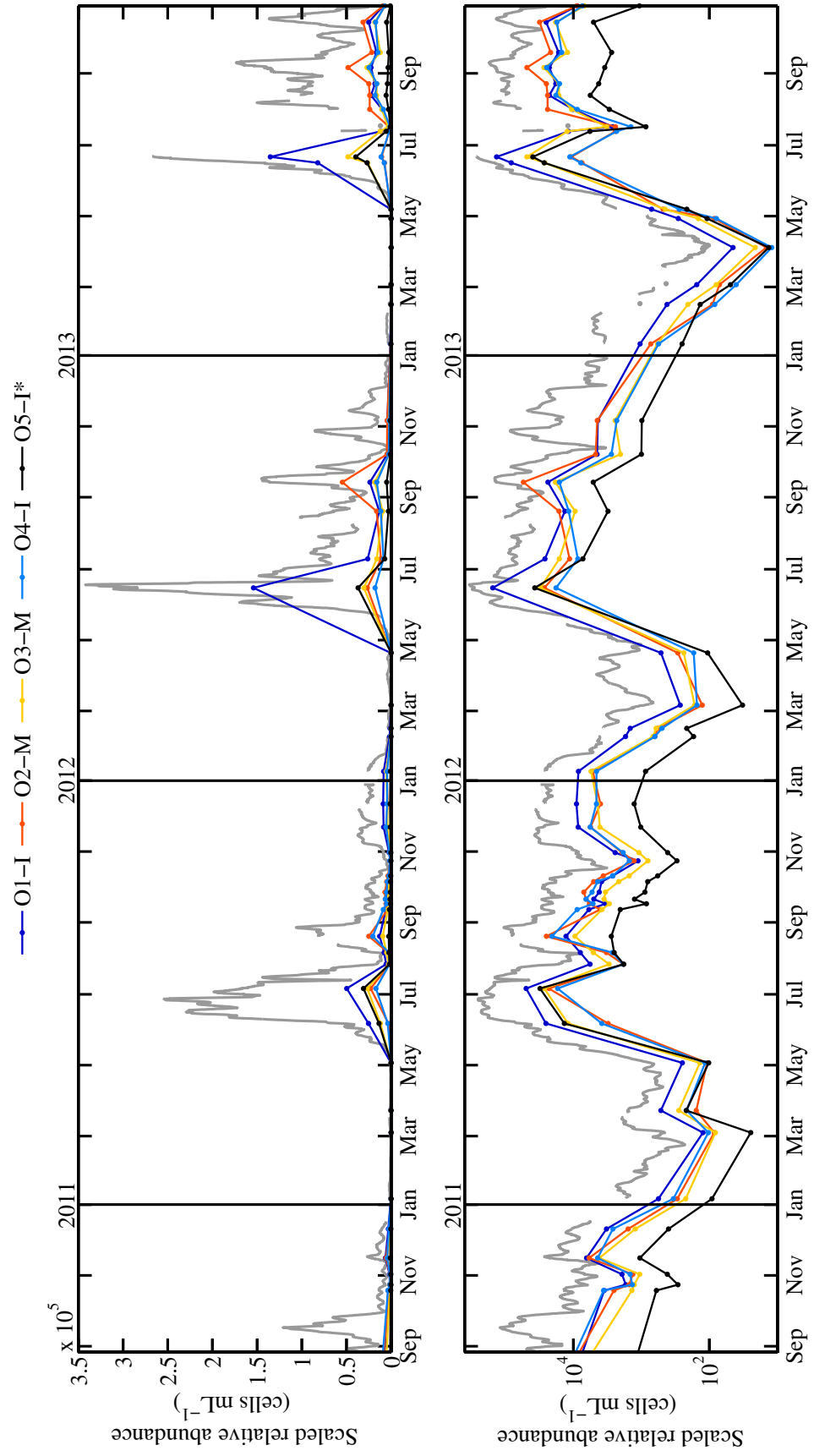
Figure 5-5: Relative abundances of each of the five most abundant tags scaled to cell concentration at the time of sampling. Top panel shows cell concentrations on linear scale, while bottom shows concentrations on a log scale. Light grey line indicates daily *Synechococcus* cell abundance. Colors are the same as in Figure 5-3 for tag legend.

137
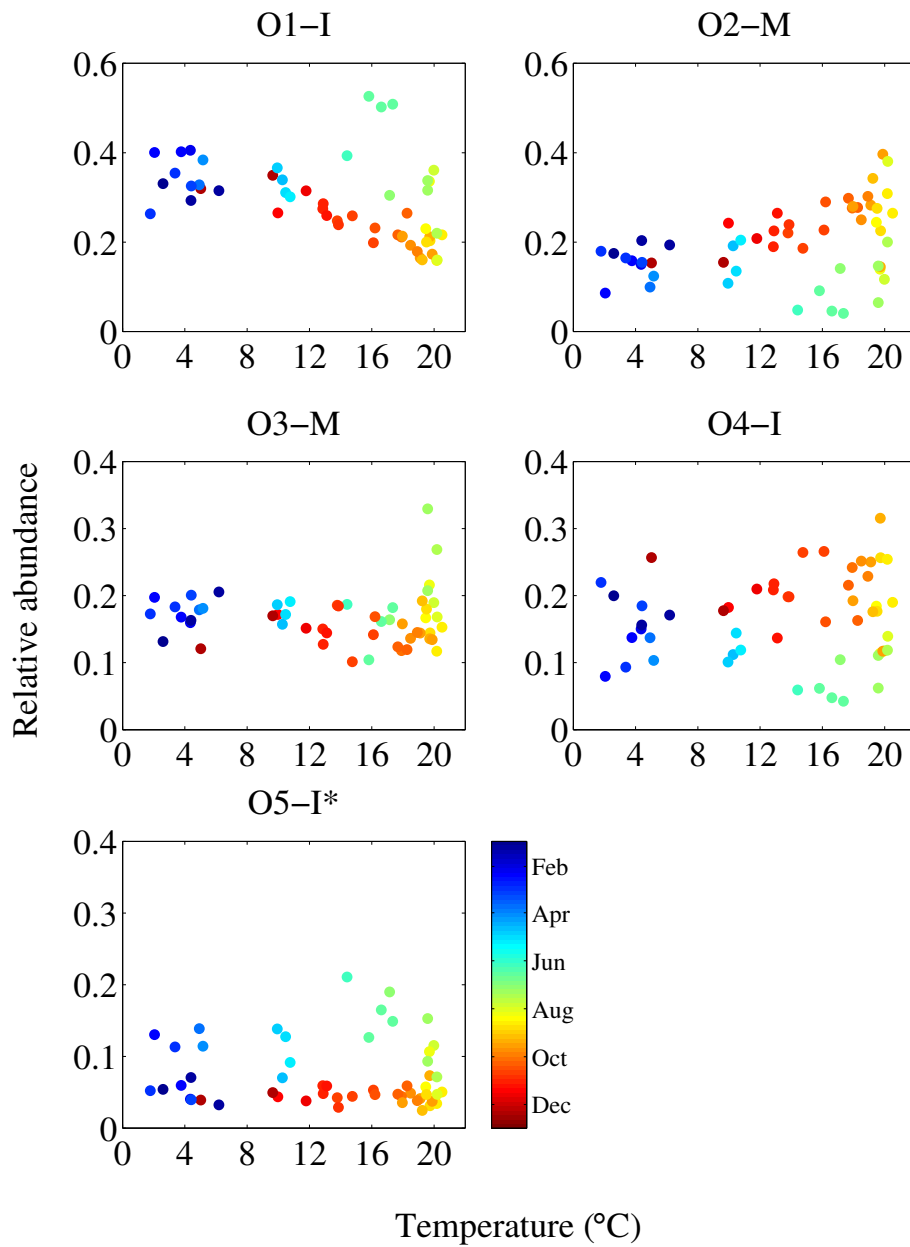
Figure 5-6: Relationship between relative abundance of the 5 most abundant tags and temperature at MVCO. Color of points indicates time of year. Note vertical scale differences between 1st row panels and the rest.
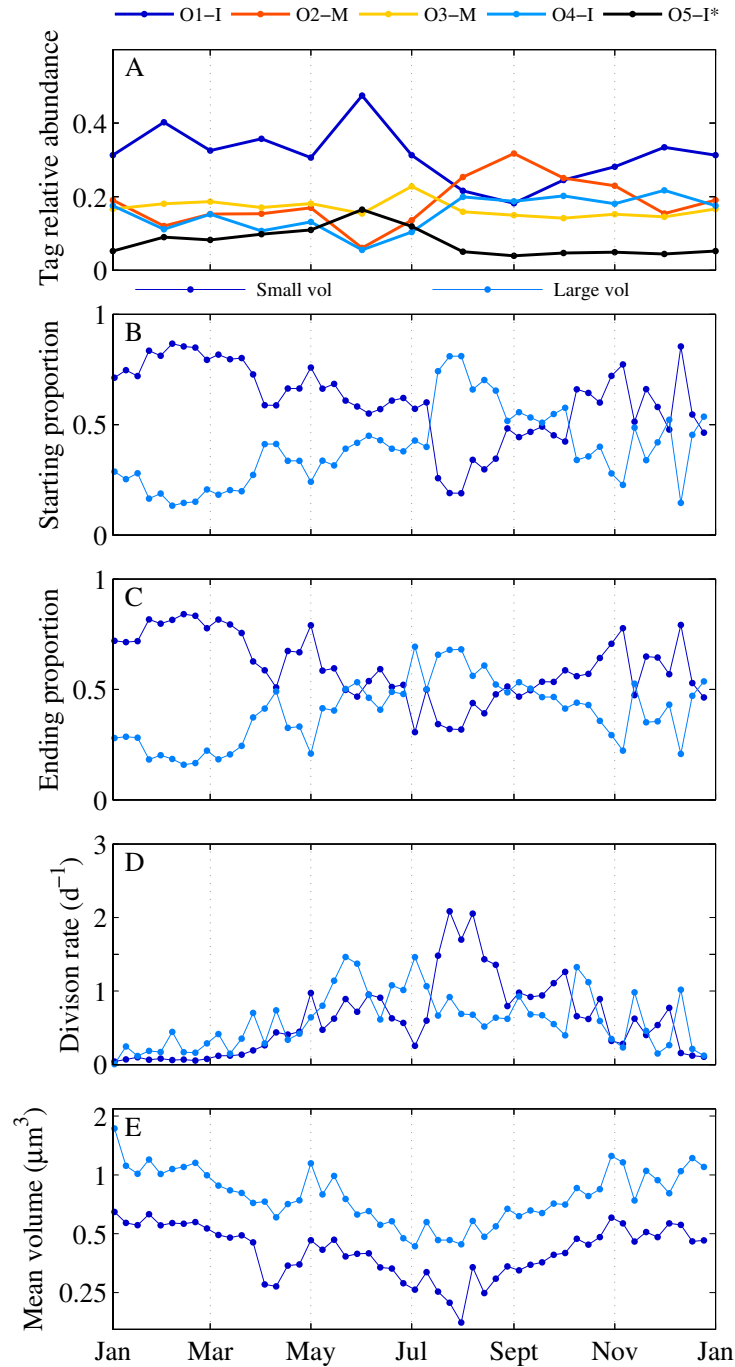
Figure 5-7: A. Monthly climatologies of relative abundance for top 5 tags. B-E. Weekly climatologies for model subpopulation characteristics. Subpopulations are distinguished by mean starting cell volume; small volume populations are denoted by dark blue, and larger volume populations denoted by light blue. The proportion of each subpopulation at the beginning hour of each day is determined by the model parameter ψ (see Chapter 2). The proportion of each subpopulation at the ending hour of the day is a function of how many cells in each subpopulation divided. B. Starting proportion of each subpopulation. C. Ending proportion of each subpopulation. D. Subpopulation division rate. E. Starting cell volumes for each subpopulation.

Table 5.1: *Synechococcus* tags for which at least 200 reads were found in the environmental sequences. Also included are tags that matched to strains isolated from MVCO that did not have at least 200 reads. Tag names are indicated in column 1 with color matching those in Figure 5-1. Column 2 denotes number of tags found within the dataset. Columns 3 and 4 indicate clade designation of sequence matches to strains isolated from MVCO and reference strains from database, respectively. Column 5 is closest strain match based on distance metrics for tags that did not have a direct match, indicated with an asterisk in column 1.

Tag name	Total	Tag match to MVCO strain	Tag match to reference strain	Closest Tag
O1-I	68,245	IC , IE	I	
O2-M	59,156	CB5, II, VII	CB5, CRD1, II, III, IV, VII, WPC2	
O3-M	44,328	CB5	II , XV	
O4-I	41,698	IE	I	
O5-I*	17,644			I
O6-IX*	2,709			IX
O7-I*	2,114			I
O8-VI	1,752	VI	VI	
O9-M	393	II	II , VI	
O10-5.2	378	5.2MV2, 5.2MV3		
O11-IX	369	IX		
O12-I*	248			I
O13-M	243	5.3I	5.3	
O14-I*	218			I
O15-VII	213		VII	
O16-CB4	196	CB4		
O17-M	100	VIII	VIII, XVI	
-	22	5.2MV1		
-	1	5.2MV1		

Chapter 6

Summary and conclusions

6.1 Thesis summary

In this thesis, I have focused on understanding the population dynamics of *Synechococcus* on the New England Shelf and how those dynamics depend on the underlying diversity structure of this important picophyotplankter. Hourly measurements of *Synechococcus* cell abundance at the Martha's Vineyard Coastal Observatory (MVCO) (from the automated flow cytometer, FlowCytobot) since 2003 reveal striking repeatable seasonal patterns, but also significant variability both on interannual and short time scales. Understanding of these patterns and cell abundance changes requires information on division rate at the same time scale. In Chapter 2, I refined and validated a matrix population model that estimates a daily population division rate from the diel changes in cell size distributions. Through experiments with both laboratory cultures and natural *Synechococcus* populations, I demonstrated that this approach accurately estimates division rate. The model provides an extremely valuable tool to be able to obtain in situ estimates of division rate at the appropriate time scale (daily) for extended periods of time (months-years).

In Chapter 3, I describe how I applied this model to the entire time series of cell size distributions obtained from FlowCytobot at MVCO. This results in an unprecedented time series of division rates. My analysis of these estimates, together with environmental data and cell abundance data, enabled insights of the controls on population dynamics. I found evidence that division rate is temperature limited during the first half of the year (January - June), but light limited during late fall and early winter (October - December). Moderate division rates in summer suggest another environmental factor (e.g., nutrient availability) is limiting division rate during this time. The loss rates that I calculated (from division rates and net growth rate)

closely track division rate in magnitude over the course of the year. Loss rates also show relationships with temperature and cell abundance, and suggest that most of the losses are due to biological agents (e.g., heterotrophic grazers and viruses). The seasonal changes in division rate coupled with almost matching loss rates determine the annual cell abundance pattern. The large seasonal changes in cell abundance from winter time lows of a few hundred cells mL^{-1} to summer time highs of 10^5 cells mL^{-1} are produced by periods when net growth rate only slightly, but systematically, deviates from zero.

Using this underlying framework of how environmental factors affect *Synechococcus* population and division rate, I am able to explain some of the observed interannual variation. The temperature of the waters surrounding MVCO has, on average, been increasing over the past decade, and we find concomitant shifts in the timing of the spring bloom. I found that *Synechococcus* division rate, as well as population abundance, begins to increase earlier in the year. It is unclear how increasing temperature will affect the dynamics for the rest of the annual cycle. Given the tight coupling of loss rate to division rate, any increase in growth of the *Synechococcus* population may just as quickly be consumed; increasing temperatures could simply result in a microbial loop that spins faster.

A key component in the interpretation of the population dynamics summarized above is knowledge of the underlying diversity structure of the *Synechococcus* assemblage. Marine *Synechococcus* are partitioned into 20 different clades, with each clade thought to occupy a distinct ecological niche. *Synechococcus* have been shown to differ in their response to environmental factors (i.e., temperature, light, nutrient availability) and hence the response of the population on the whole can be determined by the composition of the *Synechococcus* assemblage. To understand how the underlying diversity may affect the overall population dynamics, in Chapter 4, I conducted an initial survey of the diversity present year round at MVCO. With culture-dependent and independent methods, I showed that the *Synechococcus* assemblage at MVCO is indeed diverse with 13 different clades identified either by culture isolate or environmental sequences. Despite the large amount of diversity, the population seems to be dominated by clade I representatives, with apparent seasonal differences in relative abundance of subclades of clade I. I observed other clades (II, III, IV, VII, etc.) only during the late summer and fall, suggesting preferred growth conditions during this time. Flow cytometry analysis suggest that *Synechococus* that only contain phycocyanin as their primary light harvesting pigment are rare in the environment, despite a high isolation frequency.

While not strictly quantitative, the results presented in Chapter 4 suggest differences in seasonal abundance patterns of *Synechococcus* clades and subclades. I further investigated these tentative patterns in Chapter 5 with the more quantitative approach of high throughput sequencing of the 16S rRNA hypervariable V6 region of the entire bacterial community. From a three year time series of bimonthly-to-monthly samples, I found that the *Synechococcus* population is composed mainly of 5 main *oligotypes* that each demonstrate a repeatable seasonal pattern. Three oligotypes (1st, 3rd and 5th most abundant) peak during the spring bloom, while the other two oligotypes (2nd and 4th most abundant) peak during summer and may be responsible for the sustained high cell abundance and division rates observed during this time. Temperature may be an important factor that determines the relative abundance of these oligotypes, but other factors, such as cell selective grazing and targeted viral lysis may also play a role. The patterns of relative abundance coupled with the observations of cell abundance and division rate strongly suggest that population dynamics are affected by the underlying diversity structure.

6.2   Conclusions and future directions

Knowledge of a system can hinge not only on measurement and observational capability but also on modeling and innovation. High resolution observations of the *Synechococcus* population at MVCO, coupled to daily division rate estimates has enabled a much greater understanding of the population dynamics. The annual cell abundance pattern is produced by seasonal changes in division rate in response to changing environmental conditions and losses that closely match division rate in magnitude. The interplay between growth and loss processes appears to be complex, but demonstrates a systematic balance between net growth and net loss over the year. These dynamics also appear to be strongly affected by the underlying diversity structure of the *Synechococcus* population. Certain types appear to compose different seasonal features of cell abundance, which suggests that the response of the whole population is determined by the responses of individual types present.

The realization that the *Synechococcus* population is diverse at MVCO, and that certain types demonstrate different seasonal relative abundance patterns, highlights the need for knowledge about physiological and ecological differences among clade representatives. I have isolated representatives of twelve different clades from MVCO, and going forward it will be exciting to explore the physiological and ecological differences among them. Diversity at the subclade level also appears to be important.

Not all clade representatives may demonstrate the same response to environmental factors, and differences within a clade should also be explored.

The high resolution data also highlights aspects of the system that we do not yet understand. Variation in cell abundance at shorter times scales (days - weeks) should be investigated. How this short term variation affects the evolution of seasonal patterns or even long term patterns is an important question. Loss processes are a key component to this understanding, but we do not yet know how loss is partitioned among heterotrophic grazers, viral lysis and advection. On short timescales (hours - days), advection is likely to be a significant factor, but at the seasonal time scale, the main losses are likely from heterotrophic grazers or viral lysis. It will be important to partition losses between these two as the carbon fixed by *Synechococcus* can have different fates depending on the loss agent. Carbon ingested by grazers has the potential to be transferred to higher trophic levels, but the carbon released from viral lysis will be remineralized by heterotrophic bacteria.

As with any research endeavor, any one answer to a question opens up many new and exciting questions. The results presented in this thesis provide a starting point to further explore the complex and intriguing dynamics of *Synechococcus* at MVCO. Critical in this exploration will be the instruments and methodologies that enable observation of the system at the necessary time and space scales, but also the ability to leverage the unique skills, tools and capabilities found in separate disciplines. This will include endeavors in molecular biology, physical oceanography, ecological microbiology, engineering and more. A complete understanding of the dynamics will require multidisciplinary investigations into the different facets of *Synechococcus* ecology, how each piece is intertwined, and how each contributes to patterns of cell abundance.

Appendix

A.1 Second derivative finite difference calculation

Second derivatives were calculated using finite difference calculations. We used the center difference rule when both maximum likelihood estimates (MLE) of the parameters were away from the bounds:

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_j^2} = \frac{\log L(\hat{\theta}_j + h_j, \hat{\boldsymbol{\theta}}) - 2 \log L(\hat{\boldsymbol{\theta}}) + \log L(\hat{\theta}_j - h_j, \hat{\boldsymbol{\theta}})}{h_j^2} \tag{A.1}$$

and

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = \frac{1}{4 h_i h_j} [\log L(\hat{\theta}_i + h_i, \hat{\theta}_j + h_j, \hat{\boldsymbol{\theta}}) - \log L(\hat{\theta}_i + h_i, \hat{\theta}_j - h_j, \hat{\boldsymbol{\theta}})$$
$$- \log L(\hat{\theta}_i - h_i, \hat{\theta}_j + h_j, \hat{\boldsymbol{\theta}}) + \log L(\hat{\theta}_i - h_i, \hat{\theta}_j - h_j, \hat{\boldsymbol{\theta}})] \tag{A.2}$$

Here h_j is the step size chosen to approximate the derivative. For some cases, the best fit estimates of one or more parameters came back on the bound of that parameter. For these, we used a forward difference or backward difference equation to calculate the second derivative. In the case of mixed partial derivatives, where one or both parameters were on a bound, we used a combination of forward, backward and center difference equations depending on the value of the parameter.

One parameter:

Forward difference:

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_j^2} = \frac{\log L(\hat{\theta}_j + 2h_j, \hat{\boldsymbol{\theta}}) + \log L(\hat{\boldsymbol{\theta}}) - 2 \log L(\hat{\theta}_j + h_j, \hat{\boldsymbol{\theta}})}{h_j^2} \tag{A.3}$$

Backward difference:

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_j^2} = \frac{\log L(\hat{\theta}_j - 2h_j, \hat{\boldsymbol{\theta}}) + \log L(\hat{\boldsymbol{\theta}}) - 2 \log L(\hat{\theta}_j - h_j, \hat{\boldsymbol{\theta}})}{h_j^2} \tag{A.4}$$

Mixed parameters:

$\hat{\theta}_i$ at lower bound; $\hat{\theta}_j$ at lower bound:

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = \frac{1}{h_i h_j}[\log L(\hat{\theta}_i + h_i, \hat{\theta}_j + h_j, \hat{\boldsymbol{\theta}}) - \log L(\hat{\theta}_i, \hat{\theta}_j + h_j, \hat{\boldsymbol{\theta}})$$
$$- \log L(\hat{\theta}_i + h_i, \hat{\theta}_j, \hat{\boldsymbol{\theta}}) + \log L(\hat{\boldsymbol{\theta}})] \quad \text{(A.5)}$$

$\hat{\theta}_i$ at lower bound; $\hat{\theta}_j$ at upper bound:

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = \frac{1}{h_i h_j}[\log L(\hat{\theta}_i, \hat{\theta}_j + h_j, \hat{\boldsymbol{\theta}}) - \log L(\hat{\theta}_i - h_i, \hat{\theta}_j + h_j, \hat{\boldsymbol{\theta}})$$
$$+ \log L(\hat{\theta}_i - h_i, \hat{\theta}_j, \hat{\boldsymbol{\theta}}) - \log L(\hat{\boldsymbol{\theta}})] \quad \text{(A.6)}$$

$\hat{\theta}_i$ at lower bound; $\hat{\theta}_j$ not at bound:

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = \frac{1}{2h_i h_j}[\log L(\hat{\theta}_i + h_i, \hat{\theta}_j + h_j, \hat{\boldsymbol{\theta}}) - \log L(\hat{\theta}_i - h_i, \hat{\theta}_j + h_j, \hat{\boldsymbol{\theta}})$$
$$- \log L(\hat{\theta}_i + h_i, \hat{\theta}_j, \hat{\boldsymbol{\theta}}) + \log L(\hat{\theta}_i - h_i, \hat{\theta}_j, \hat{\boldsymbol{\theta}})] \quad \text{(A.7)}$$

$\hat{\theta}_i$ at upper bound; $\hat{\theta}_j$ at upper bound:

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = \frac{1}{h_i h_j}[\log L(\hat{\boldsymbol{\theta}}) - \log L(\hat{\theta}_i - h_i, \hat{\boldsymbol{\theta}})$$
$$- \log L(\hat{\theta}_i, \hat{\theta}_j - h_j, \hat{\boldsymbol{\theta}}) + \log L(\hat{\theta}_i - h_i, \hat{\theta}_j - h_j, \hat{\boldsymbol{\theta}})] \quad \text{(A.8)}$$

$\hat{\theta}_i$ at upper bound; $\hat{\theta}_j$ not at bound:

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = \frac{1}{2h_i h_j}[\log L(\hat{\theta}_i + h_i, \hat{\theta}_j, \hat{\boldsymbol{\theta}}) - \log L(\hat{\theta}_i - h_i, \hat{\theta}_j, \hat{\boldsymbol{\theta}})$$
$$- \log L(\hat{\theta}_i + h_i, \hat{\theta}_j - h_j, \hat{\boldsymbol{\theta}}) + \log L(\hat{\theta}_i - h_i, \hat{\theta}_j - h_j, \hat{\boldsymbol{\theta}})] \quad \text{(A.9)}$$

A.2 Model simulations and structure comparisons

Figures A-1 - A-9 are histograms of the maximum likelihood estimates (MLE) of parameters from 100 cases of simulated data from specified parameters. The parameters used to simulate each case are listed in Table A.1. For the first cases, data was simulated with 500 and 5000 cells observed in each hour, whereas the second cases were

only simulated with 5000 observed cells in an hour. In general, we find the model and solver routine are able to retrieve the parameter values that were used for simulation with sampling from a Dirichlet-multinomial distribution. The MLEs of parameters for cases simulated with more cells are, in general, closer to the parameters used to simulate the data.

Figures A-10 - A-13 illustrate how different model versions vary in their ability to estimate division rate of both laboratory cultures and natural *Synechococcus* populations. Many of the decisions for model structure stemmed from these comparisons.

Table A.1: Parameters used to simulate data. Parameters are grouped as generating either a 'low', 'medium' or 'high' value of division rate.

	Case	$\gamma_{max}1$	$b1$	E^*1	$\delta_{max}1$	$\gamma_{max}2$	$b2$	E^*2	$\delta_{max}2$	ψ	$\bar{\nu}_1$	$\bar{\nu}_2$	σ^2	s	μ
Low	1	0.06	3	200	0.02	0.04	5	600	0.005	0.3	30	36	4	600	0.15
	2	0.13	2	200	0.04	0.05	4	400	0.01	0.25	30	35	6	3000	0.41
Med.	1	0.12	4	60	0.07	0.08	3.5	200	0.1	0.4	32	28	6	4000	0.66
	2	0.09	3.5	65	0.05	0.05	3.5	175	0.03	0.8	30	37	4.2	1800	0.50
High	1	0.2	4	3	0.6	0.05	6	10	0.1	0.45	27	36	4	1300	1.09
	2	0.14	3	60	0.06	0.06	2	400	0.08	0.4	28	34	5	1500	0.97
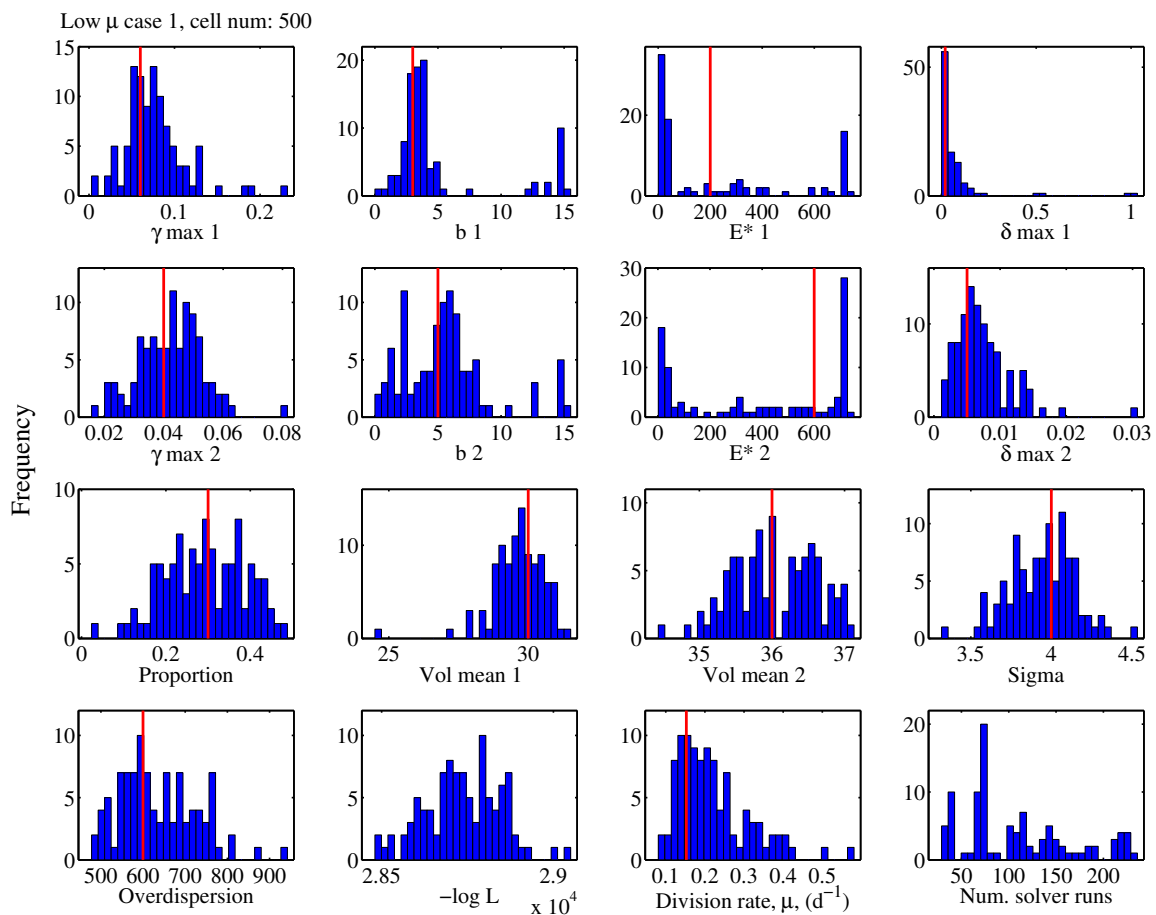
Figure A-1: Histograms of the MLEs of parameters, negative log likelihood values, calculated division rates from parameters, and number of solver runs utilized by the optimization routine to find a global minimum for low division rate case 1 with 500 cells. Parameter values used to simulate data and division rate from these parameters are indicated by vertical red lines.
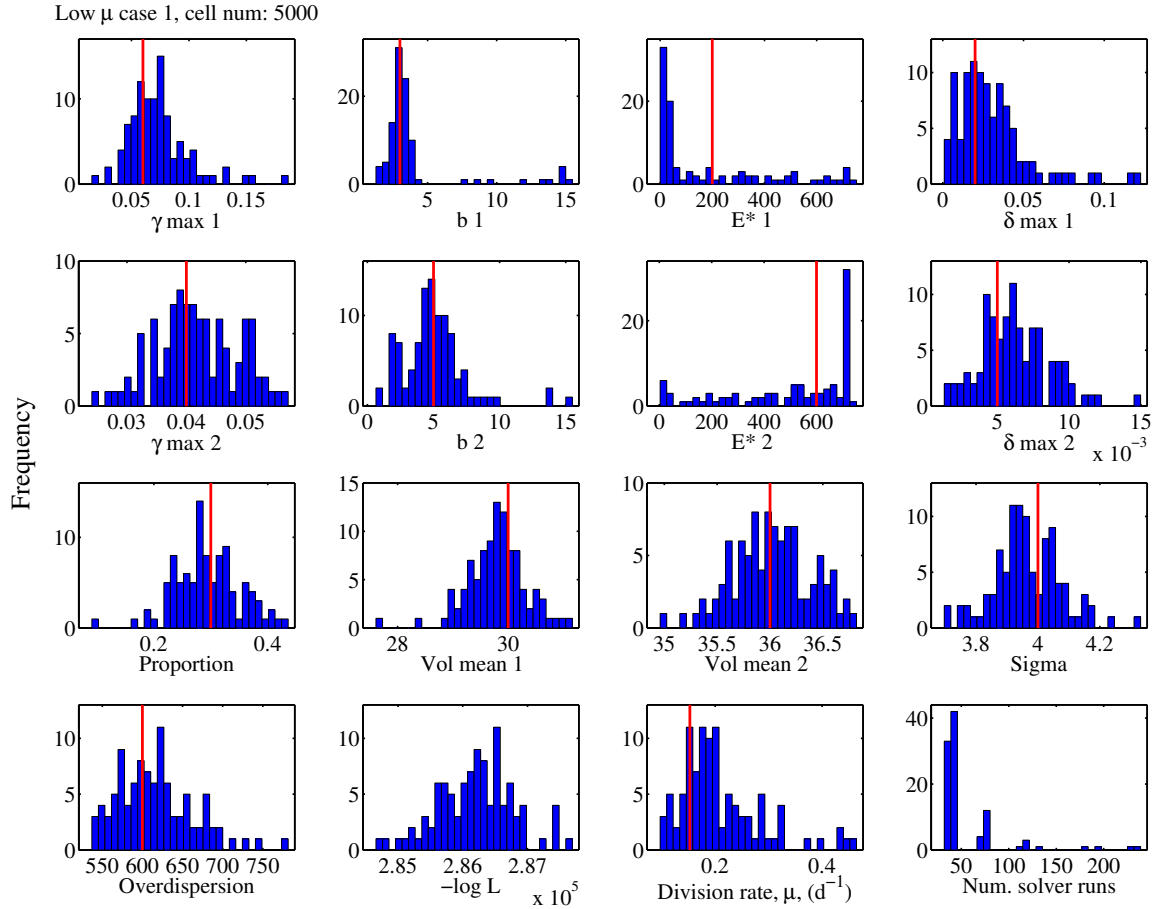
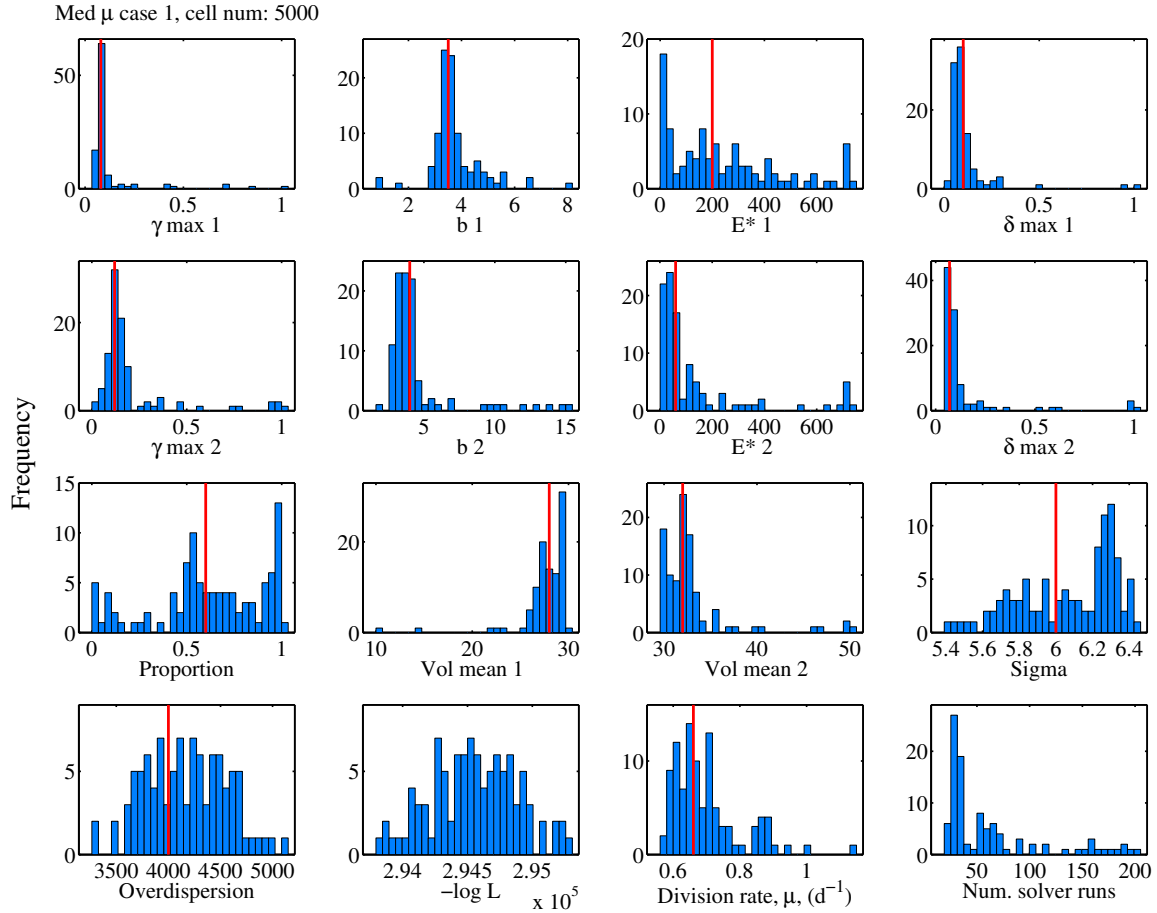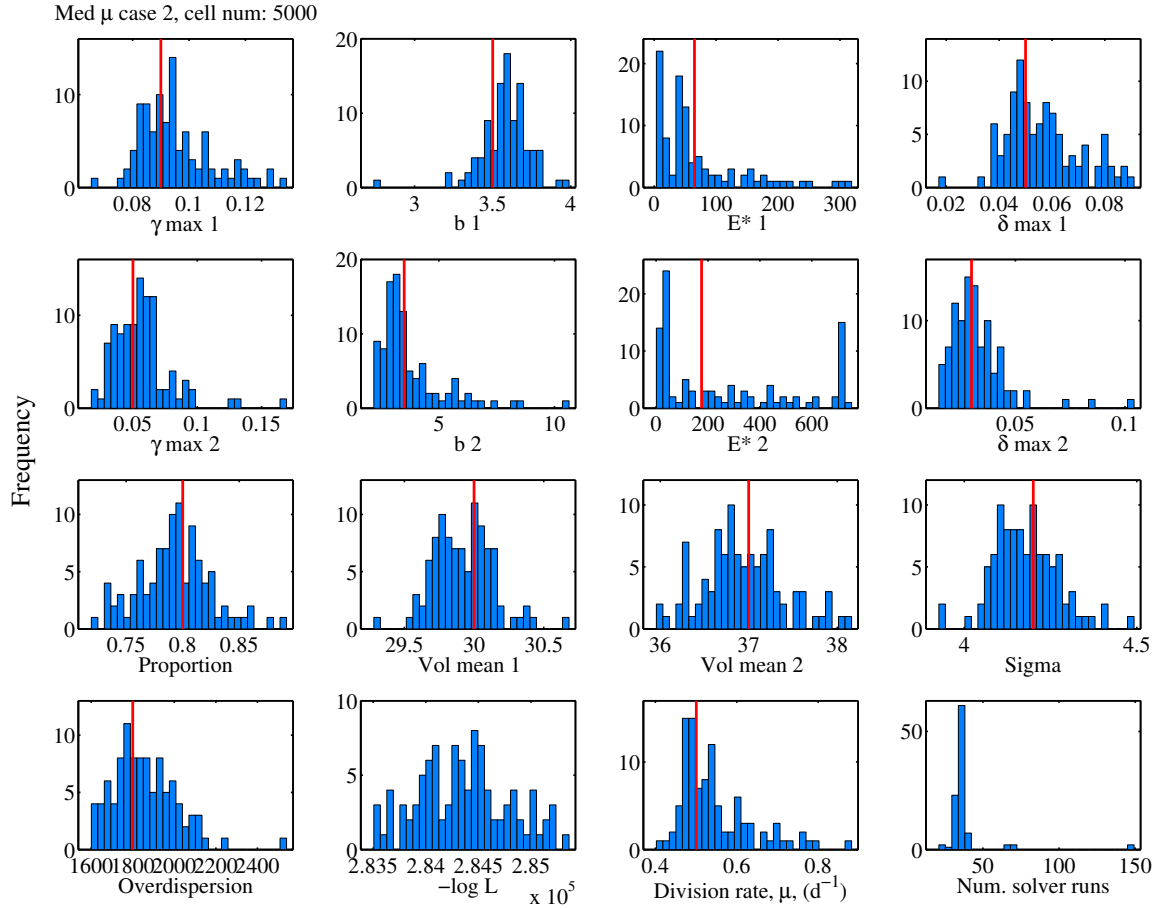Figure A-2: Histograms of the MLEs of parameters, negative log likelihood values, calculated division rates from parameters, and number of solver runs utilized by the optimization routine to find a global minimum for low division rate case 1 with 5000 cells. Parameter values used to simulate data and division rate calculated from these parameters are indicated by vertical red lines.
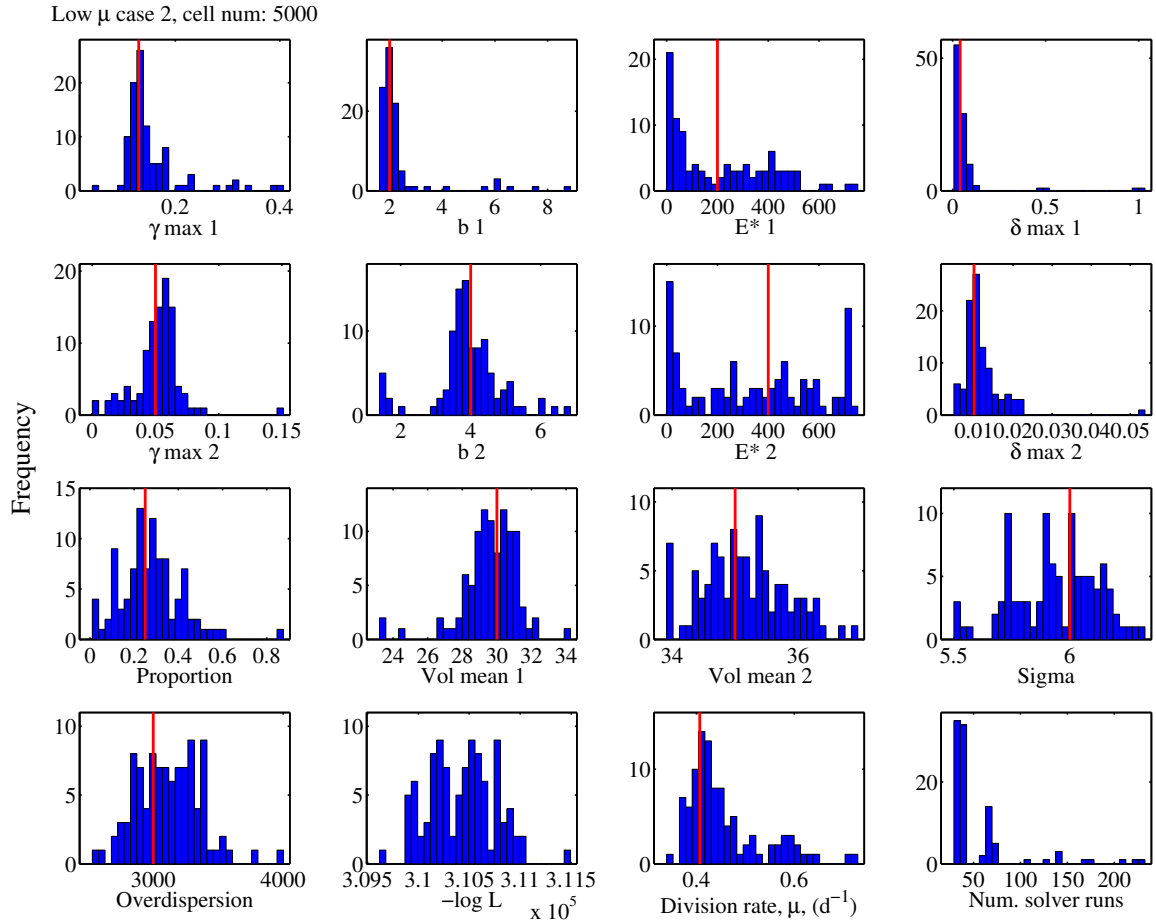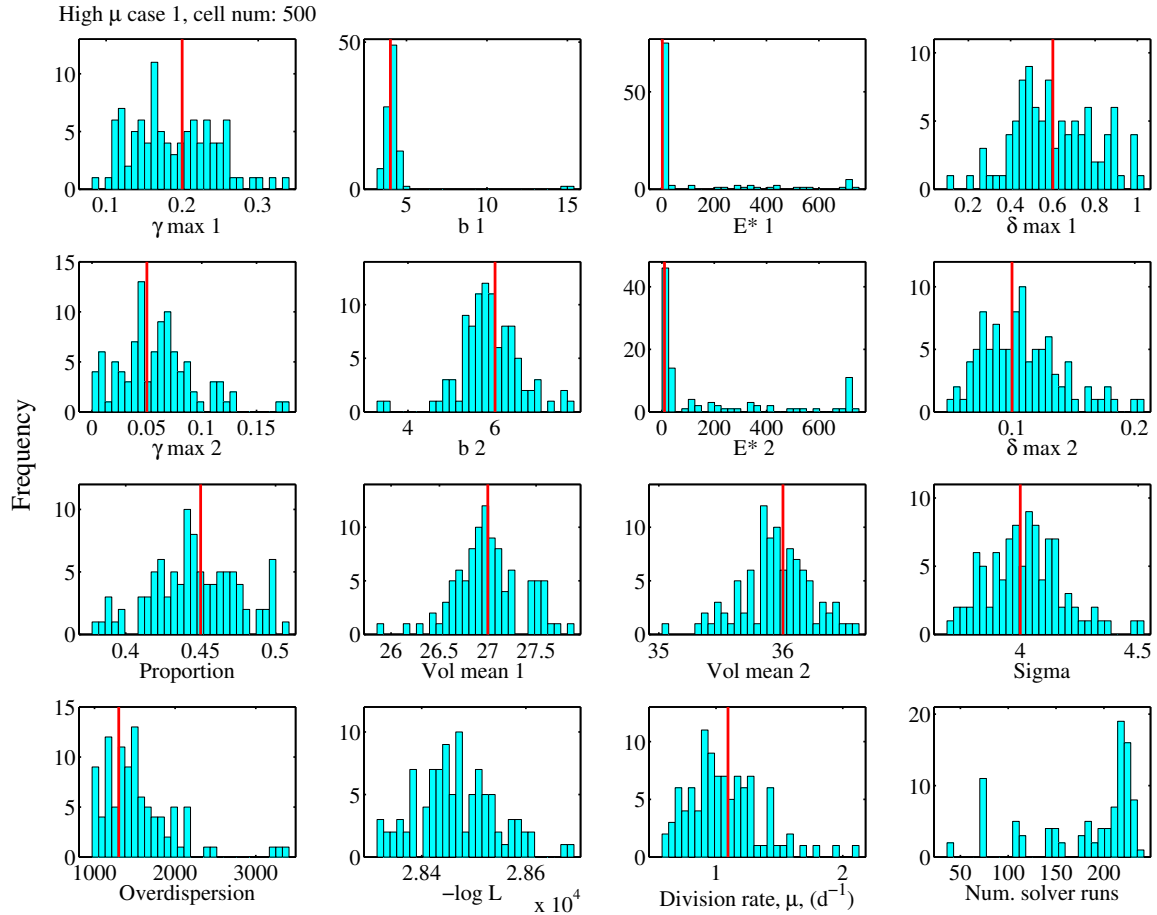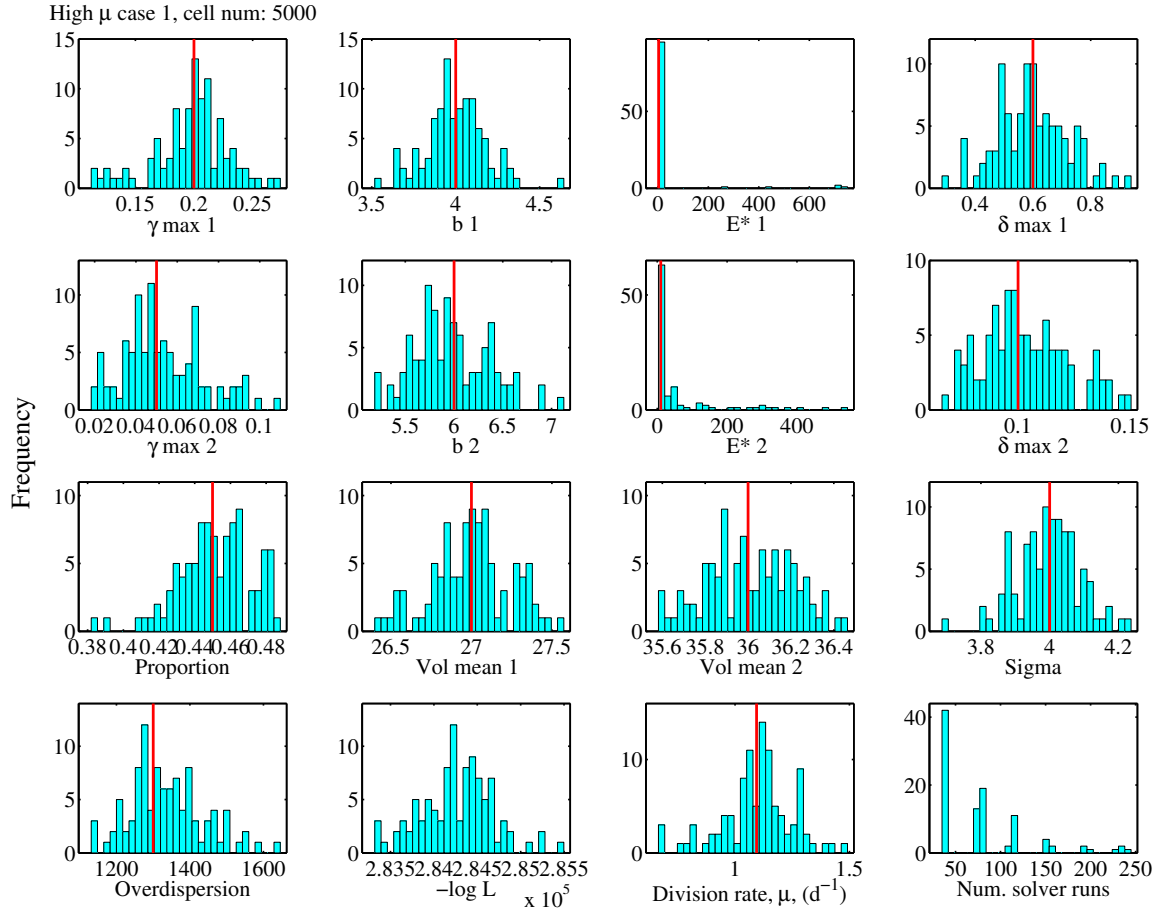
Figure A-3: Histograms of the MLEs of parameters, negative log likelihood values, calculated division rates from parameters, and number of solver runs utilized by the optimization routine to find a global minimum for low division rate case 2 with 5000 cells. Parameter values used to simulate data and division rate calculated from these parameters are indicated by vertical red lines.
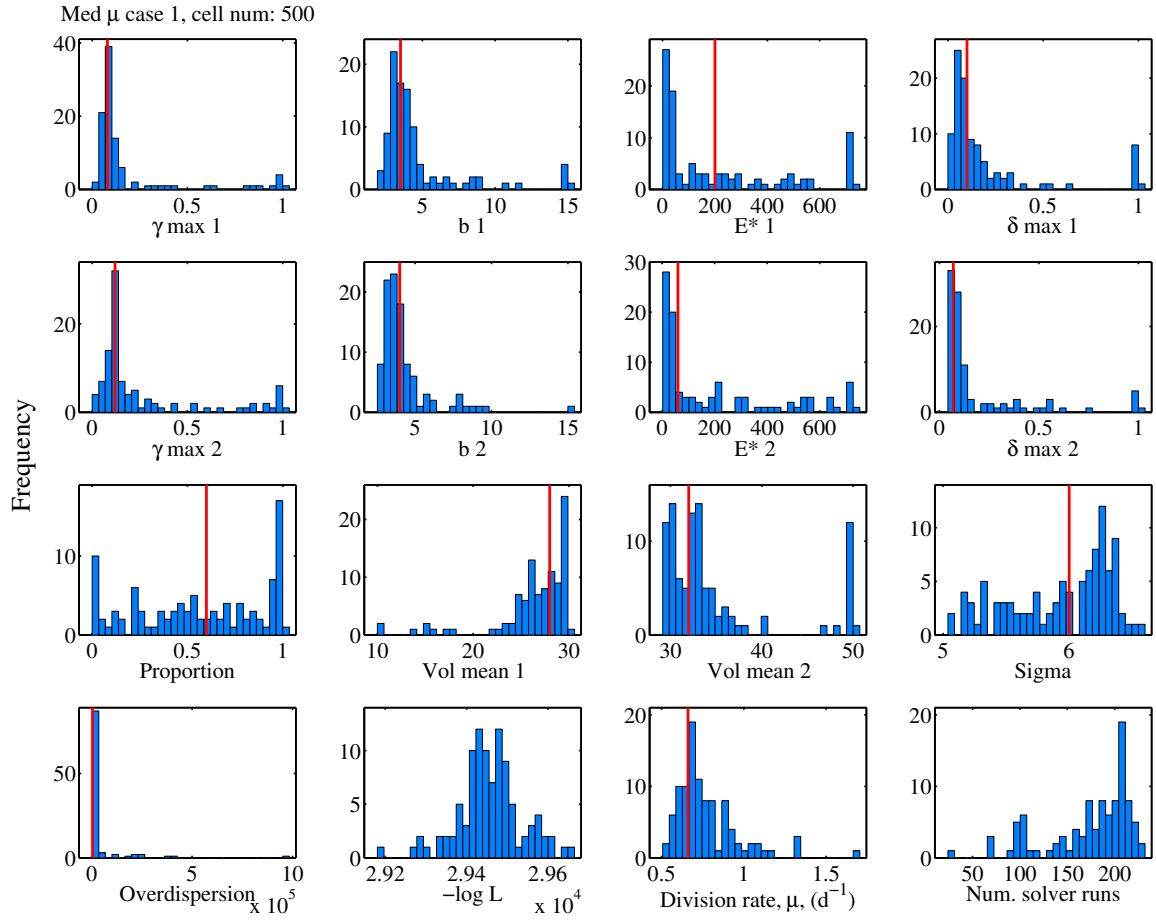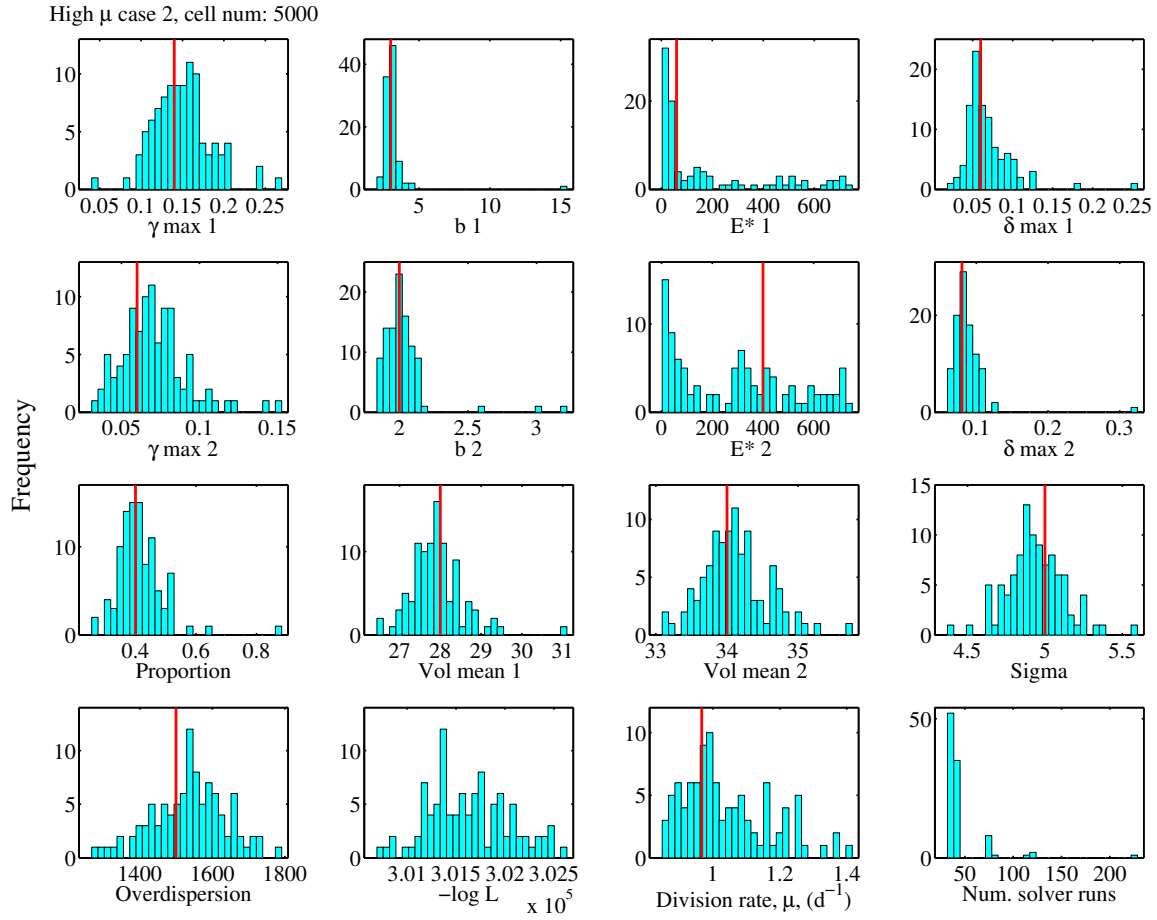
Figure A-4: Histograms of the MLEs of parameters, negative log likelihood values, calculated division rates from parameters, and number of solver runs utilized by the optimization routine to find a global minimum for medium division rate case 1 with 500 cells. Parameter values used to simulate data and division rate calculated from these parameters are indicated by vertical red lines.

Figure A-5: Histograms of MLE of parameters, negative log likelihood values, calculated division rates from parameters, and number of solver runs utilized by the optimization routine to find a global minimum or medium division rate case 1 with 5000 cells. Parameter values used to simulate data and division rate calculated from these parameters are indicated by vertical red lines.

Figure A-6: Histograms of MLE of parameters, negative log likelihood values, calculated division rates from parameters, and number of solver runs utilized by the optimization routine to find a global minimum or medium division rate case 2 with 5000 cells. Parameter values used to simulate data and division rate calculated from these parameters are indicated by vertical red lines.

Figure A-7: Histograms of MLE of parameters, negative log likelihood values, calculated division rates from parameters, and number of solver runs utilized by the optimization routine to find a global minimum or high division rate case 1 with 500 cells. Parameter values used to simulate data and division rate calculated from these parameters are indicated by vertical red lines.

Figure A-8: Histograms of MLE of parameters, negative log likelihood values, calculated division rates from parameters, and number of solver runs utilized by the optimization routine to find a global minimum or high division rate case 1 with 5000 cells. Parameter values used to simulate data and division rate calculated from these parameters are indicated by vertical red lines.

Figure A-9: Histograms of MLE of parameters, negative log likelihood values, calculated division rates from parameters, and number of solver runs utilized by the optimization routine to find a global minimum or high division rate case 2 with 5000 cells. Parameter values used to simulate data and division rate calculated from these parameters are indicated by vertical red lines.
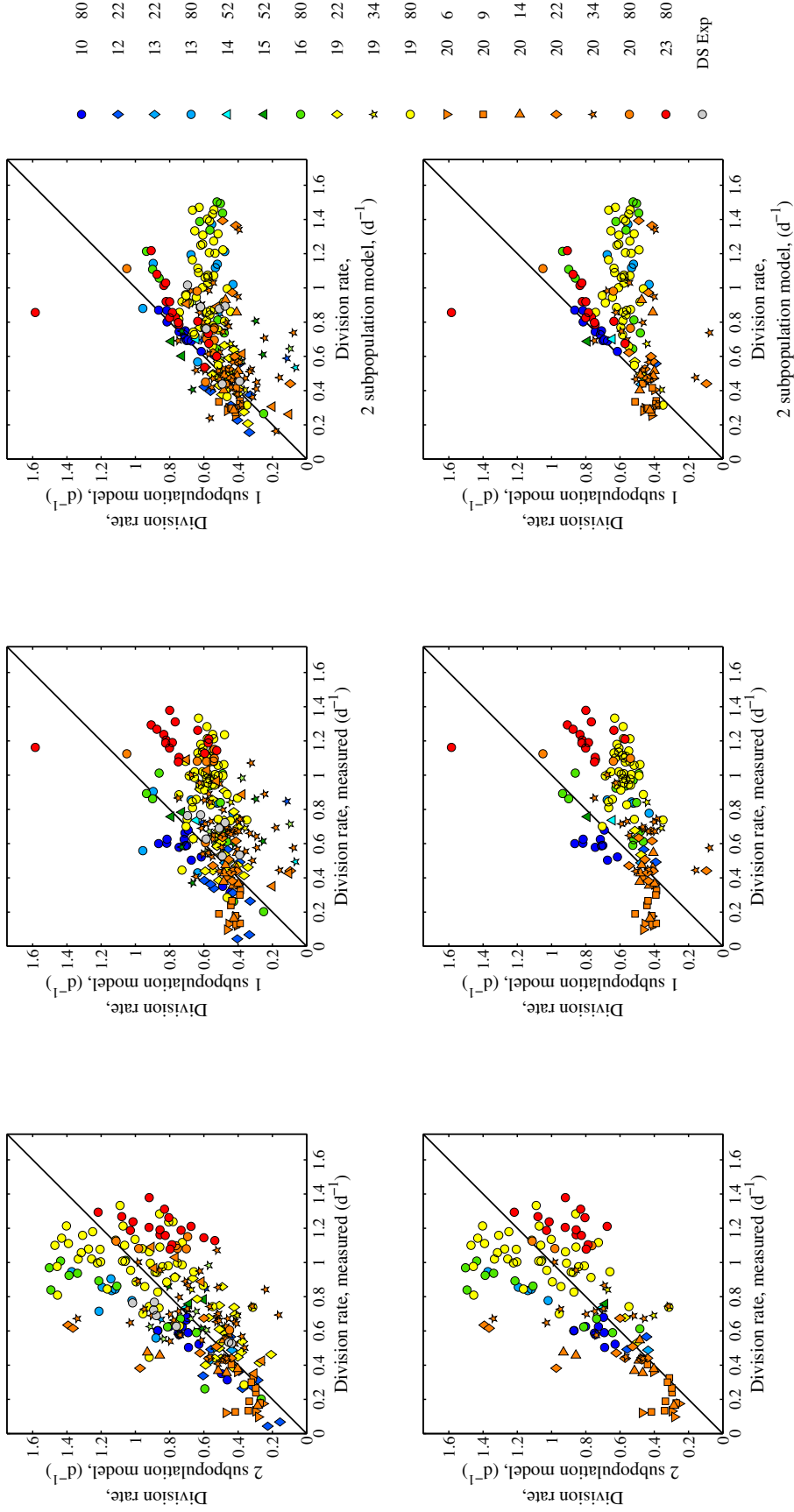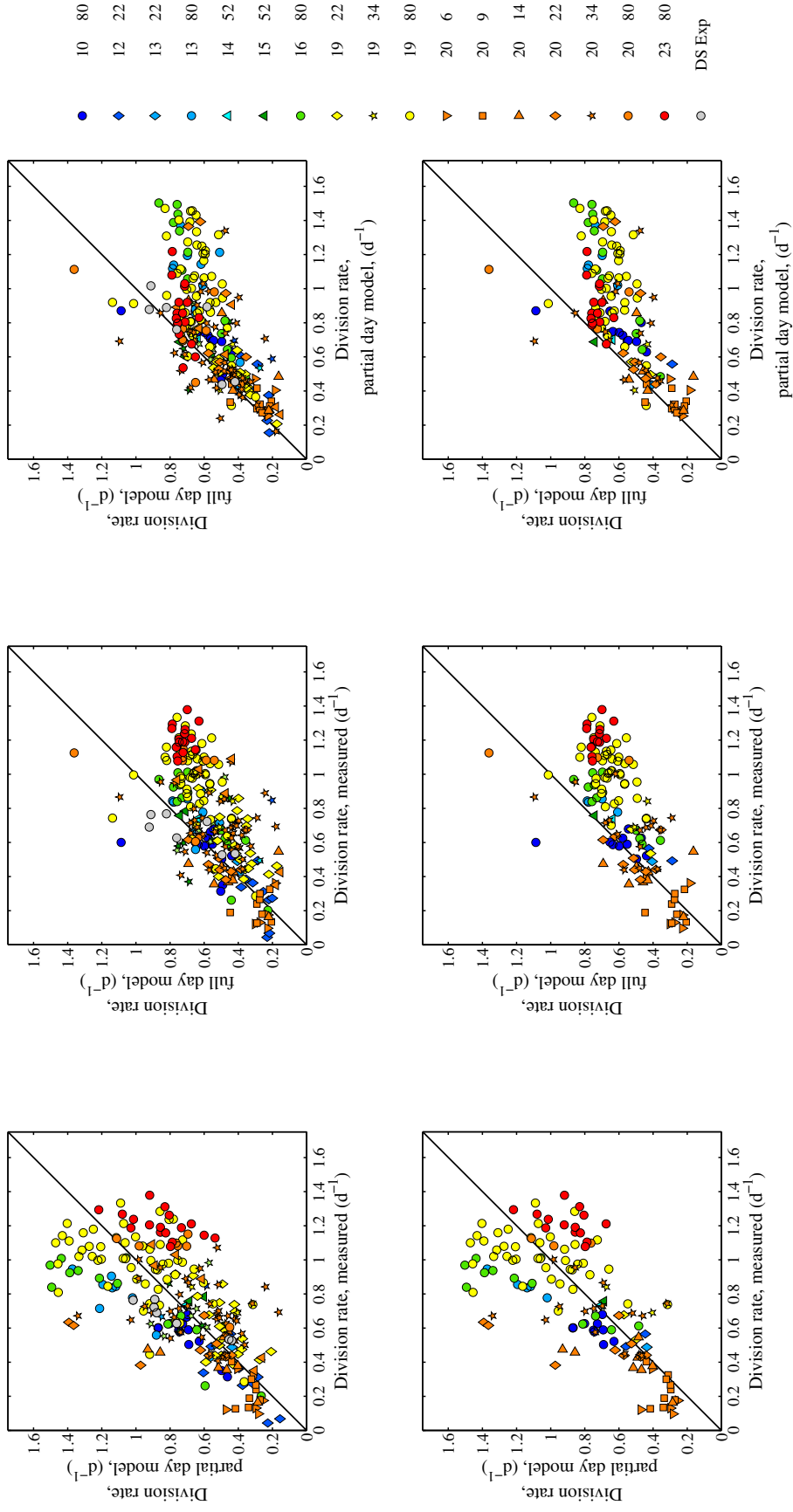
Figure A-10: Comparisons between model estimated division rate and measured division rate for both laboratory cultures (obtained from counts) and natural *Synechococcus* (obtained from the dilution series method). These comparisons illustrate how different model versions (either one population or two subpopulations) vary in their ability to estimate division rate. The first row includes all laboratory data for the final comparisons, whereas the second row only contains data that did not exhibit 'long tails' in the cell size distributions. In general, the one population model is not able to accurately estimate the division rate as well as the two subpopulation model, and further supports our decision to incorporate two populations into the model structure.
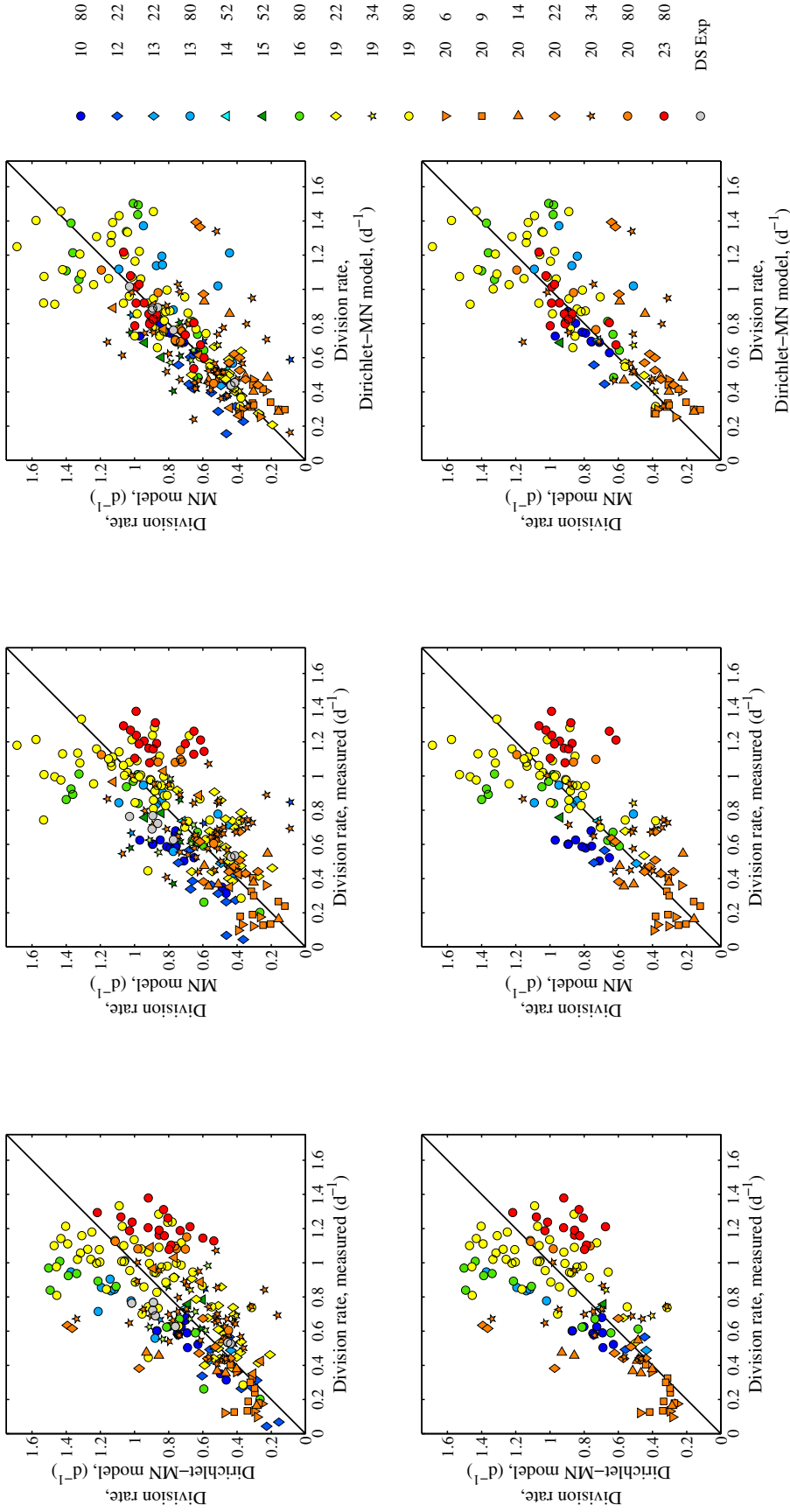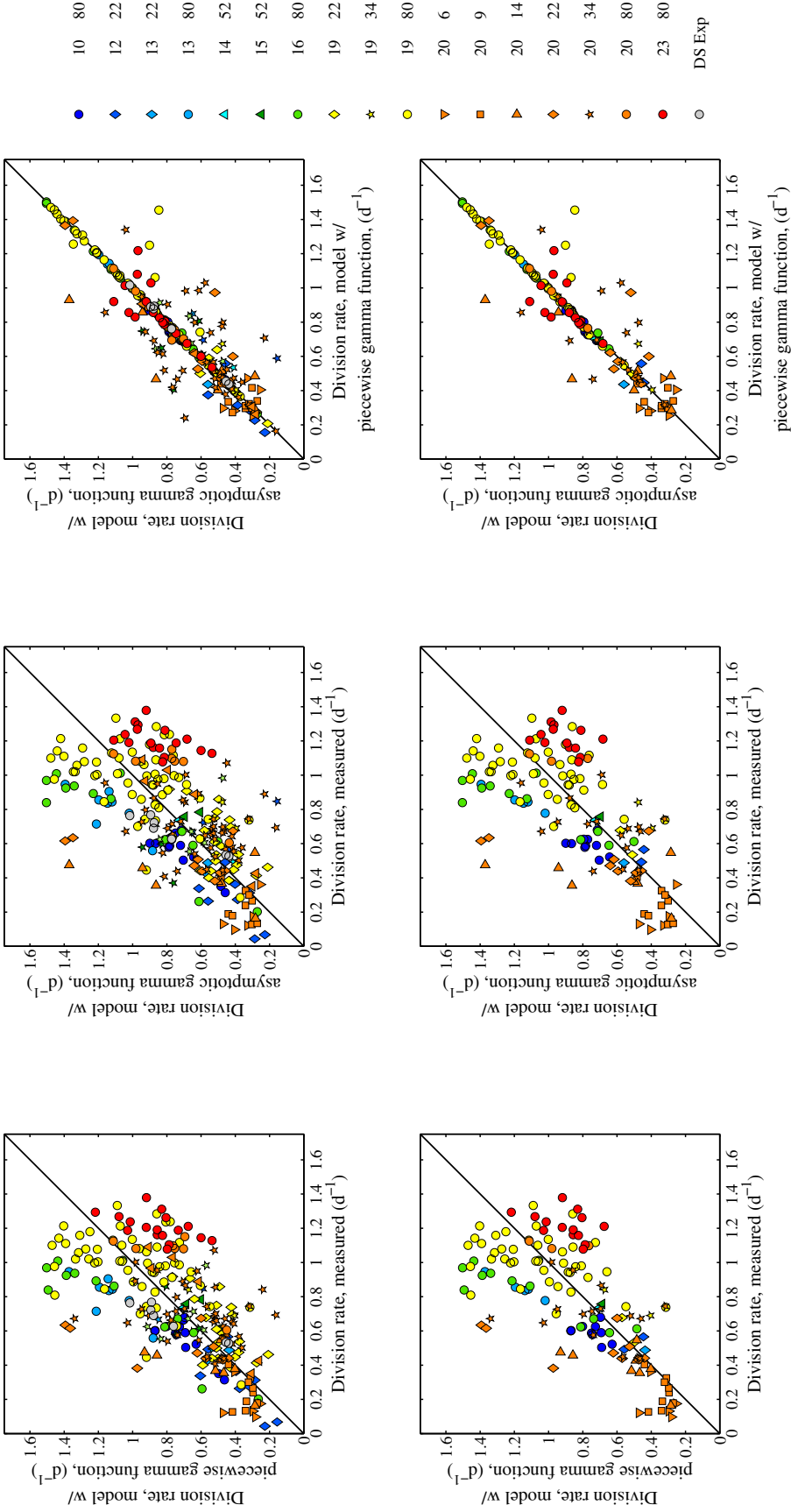
Figure A-11: Comparisons between model estimated division rate and measured division rate for both laboratory cultures (obtained from counts) and natural *Synechococcus* (obtained from the dilution series method). These comparisons illustrate the effect of different likelihood functions, with the assumption that observations result from either a multinomial or Dirichlet-multinomial sampling process, on accurate estimation of division rate. The first row includes all laboratory data for the final comparisons, whereas the second row only contains data that did not exhibit 'long tails' in the cell size distributions. While the results appear similar, in general, the choice of the Dirichlet-multinomial distribution results in a closer match to measured division rates from the laboratory cultures.

Figure A-12: Comparisons between model estimated division rate and measured division rate for both laboratory cultures (obtained from counts) and natural *Synechococcus* (obtained from the dilution series method). These comparisons illustrate the ability of either the full day model (uses all 24 hours of observations) or the a partial day model (uses hours 7-25 after dawn) to accurately estimate of division rate. The first row includes all laboratory data for the final comparisons, whereas the second row only contains data that did not exhibit 'long tails' in the cell size distributions. We find that only using a partial day results in much more accurate estimations of division rate. In particular, the full day model does not estimate division rate above 0.8 d^{-1} well. This indicates that the physiological processes that govern cell size in the first 6 hours after dawn are not represented well by our model equations. In particular, we find that there is generally little change in cell volume during these hours, suggesting a latent period before volume can increasing, despite increasing radiation during this time. Our model requires cells to increase in size with light, and this is likely the piece of the model that is responsible for the inaccurate division rates.

Figure A-13: Comparisons between model estimated division rate and measured division rate for both laboratory cultures (obtained from counts) and natural *Synechococcus* (obtained from the dilution series method). These comparisons illustrate how different growth functions, γ, effect estimation of division rate. The first row includes all laboratory data for the final comparisons, whereas the second row only contains data that did not exhibit 'long tails' in the cell size distributions. In general, division rate estimation is very similar between model versions with different gamma functions. This is not surprisingly as these two functions are very similar. The decision to use the piecewise linear function was an initial attempt to make the likelihood space less difficult for the solver to search through.

Appendix for Chapter 3
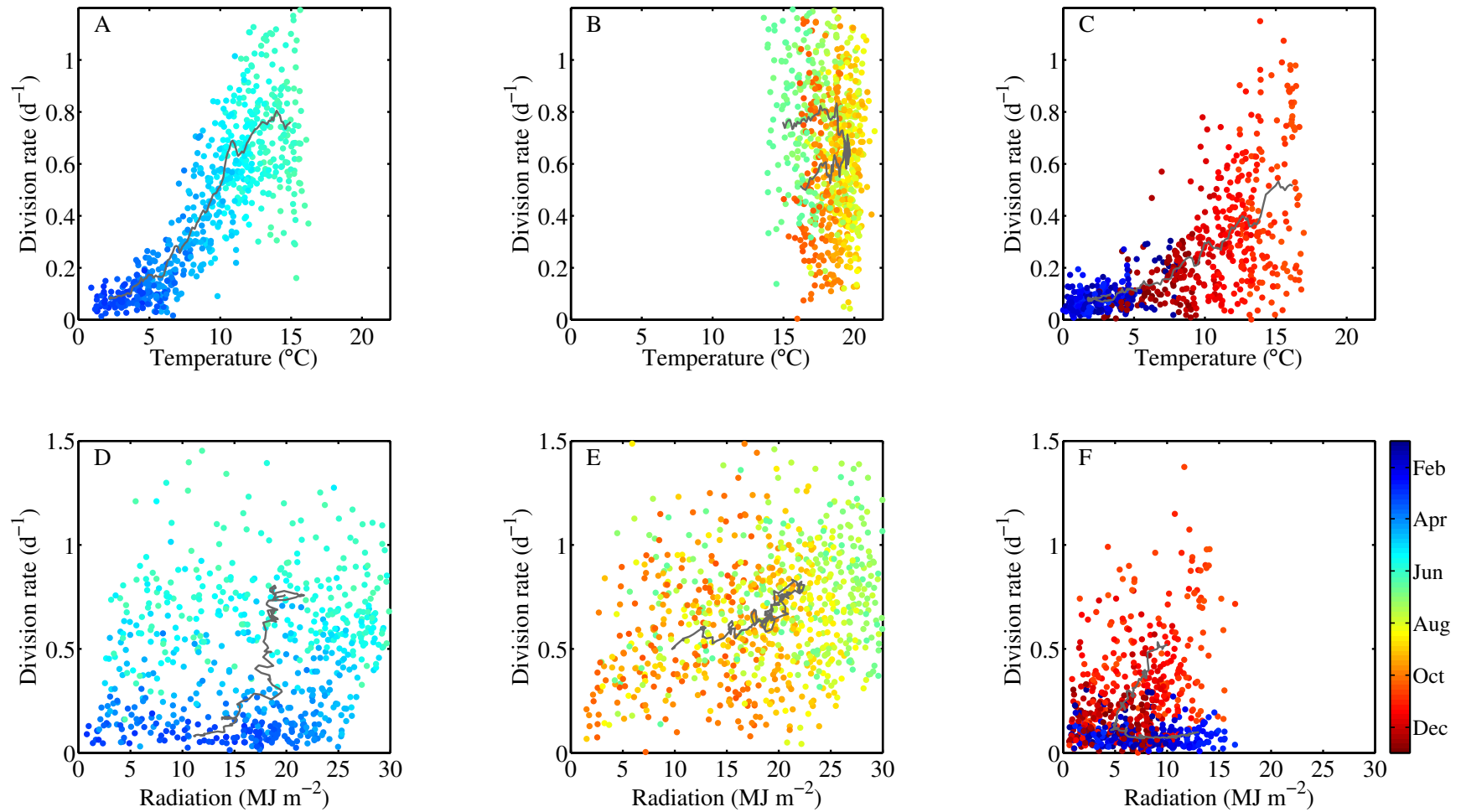
Figure A-14: Relationships between division rate and temperature for each datapoint in the time series. Panels contain points belonging to different times of year. A. March 1 - June 15. B. June 16 - October 15. C. October 15 - February 28. Relationships between division rate and light for each datapoint in the time series. Panels contain points belonging to different times of year. D. March 1 - June 15. E. June 16 - October 15. F. October 15 - February 28. Color indicates time of year.
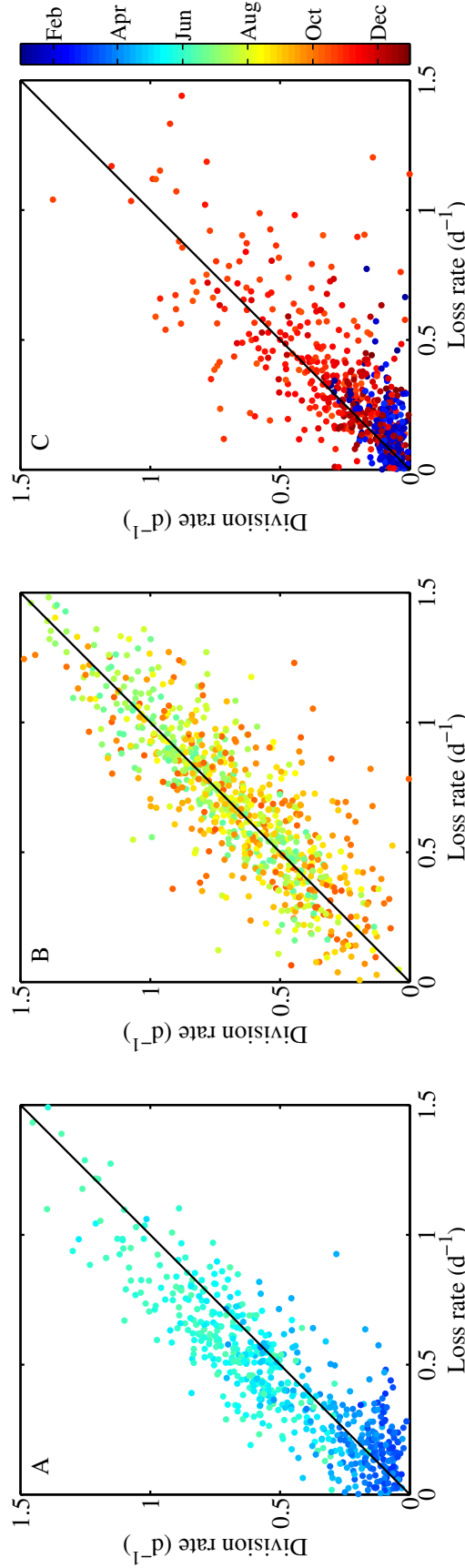
Figure A-15: Relationships between division rate and loss rate for each datapoint in the time series. Panels contain points belonging to different times of year. A. March 1 - June 15. B. June 16 - October 15. C. October 15 - February 28. Color indicates time of year.
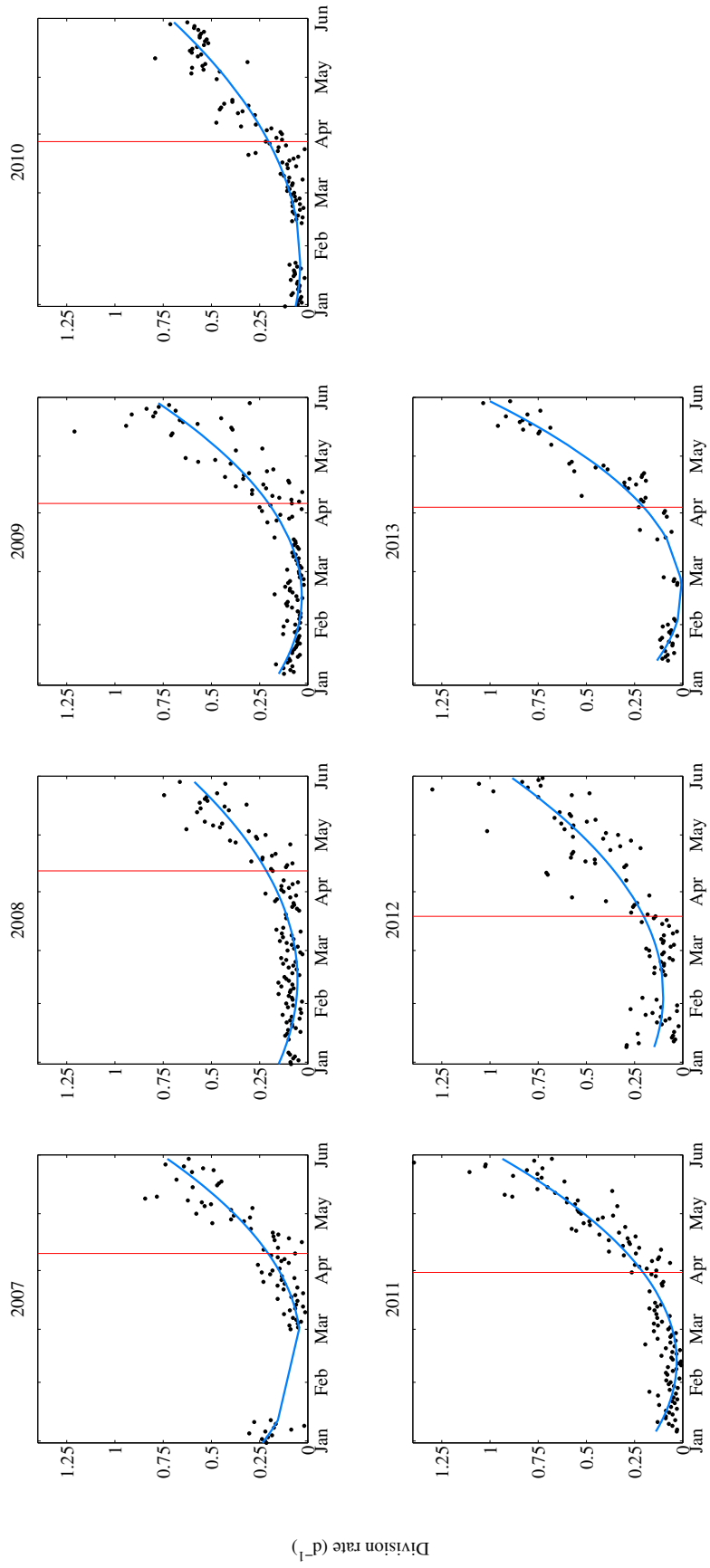
Figure A-16: Division rates estimated from model for 2007 - 2013 during the first half of the year (black markers). Blue line indicates second order polynomial fit used to determine when division rate started increasing for the season (indicated by red line).
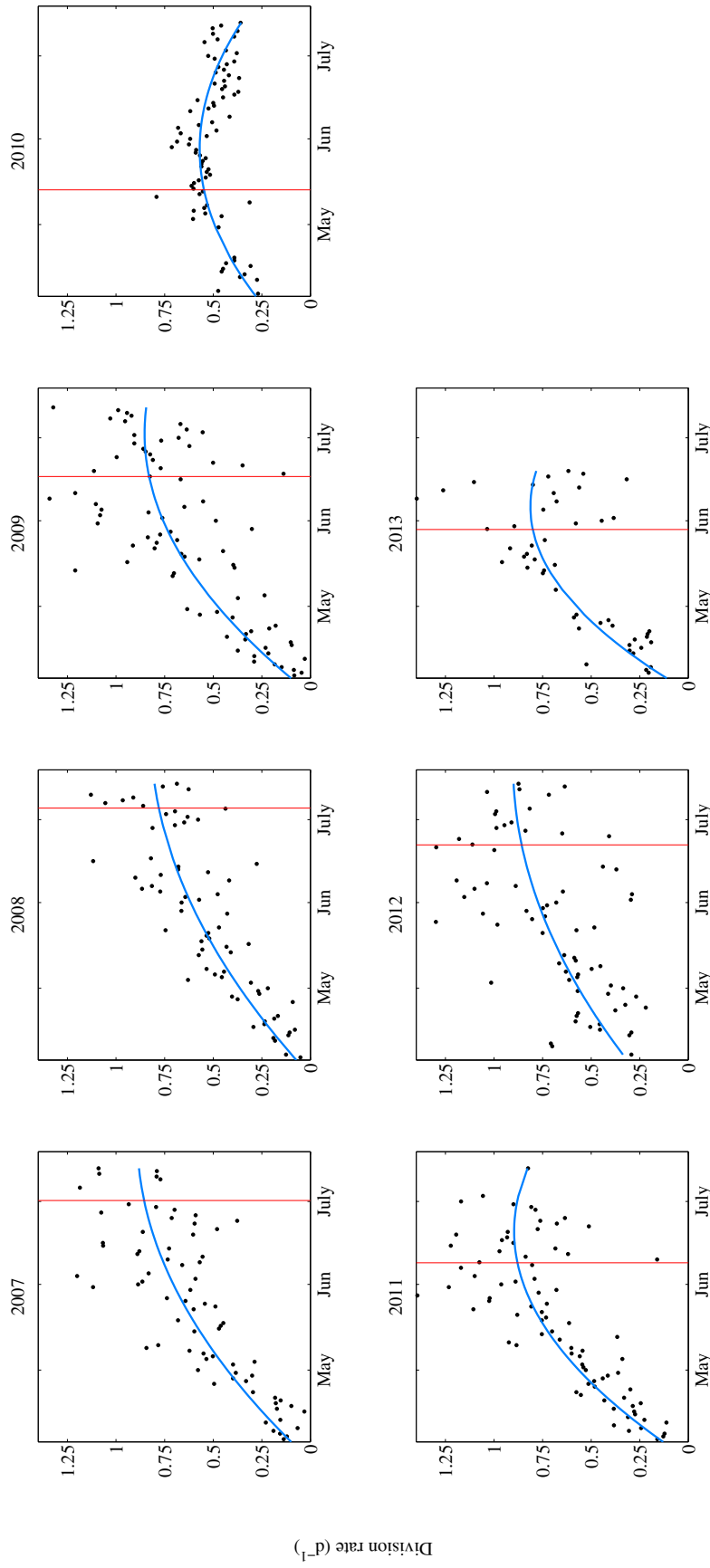
Figure A-17: Division rates estimated from model for 2007 - 2013 from April to mid-July (black markers). Blue line indicates second order polynomial fit used to determine when division rate stopped increasing for the season (indicated by red line).
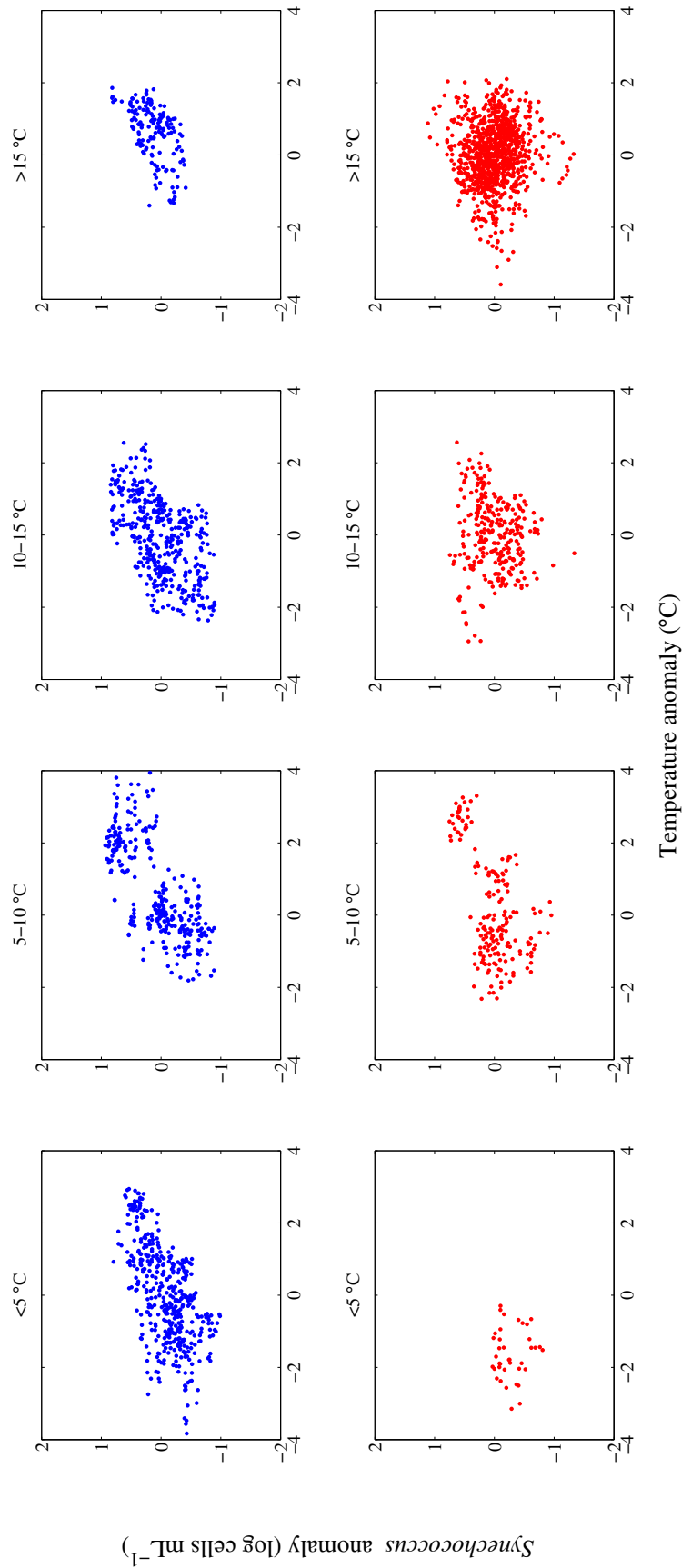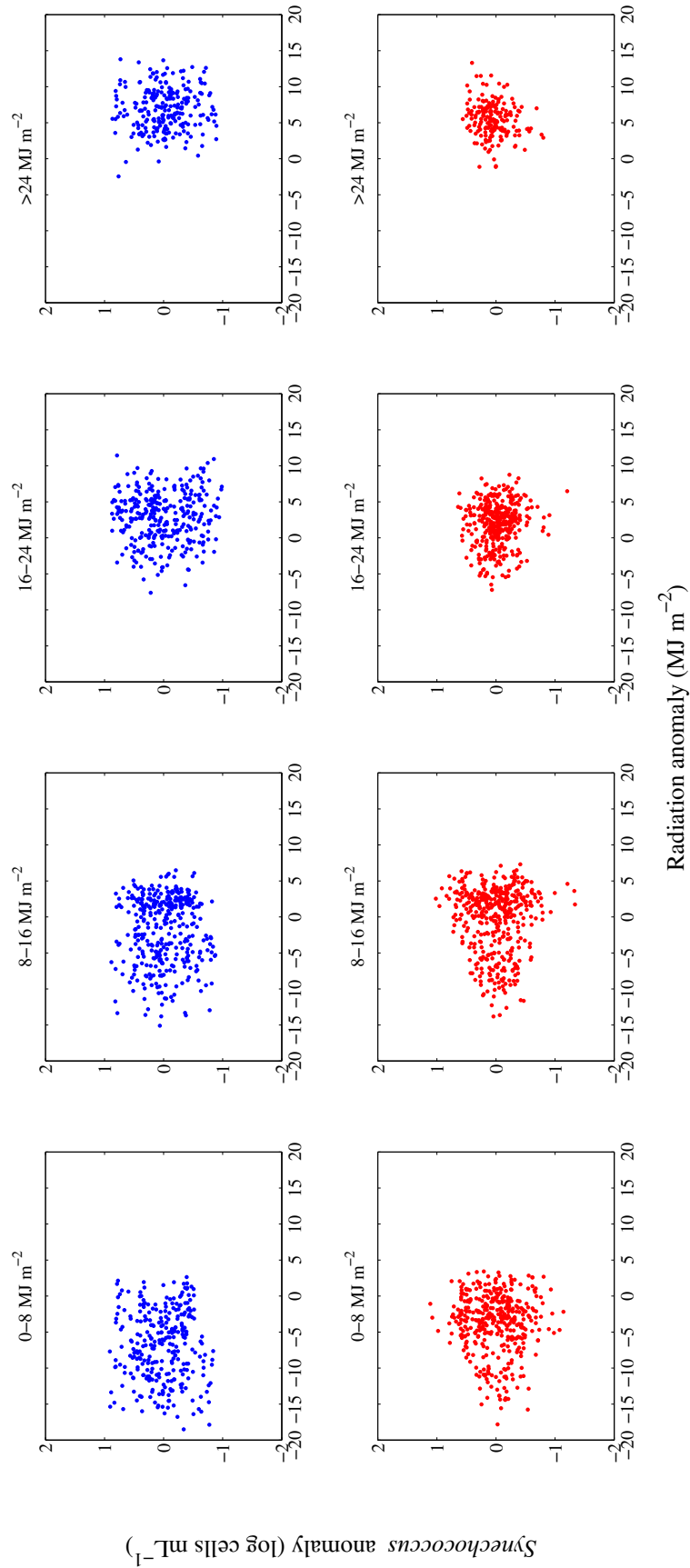
Figure A-18: The top row shows the relationships between anomalies of *Synechococcus* cell concentration and anomalies of temperature for different temperature intervals for the first half of the year (January - June) (indicated as blue points), while the bottom row shows these relationships for the second half of the year (July-December) (indicated as red points).
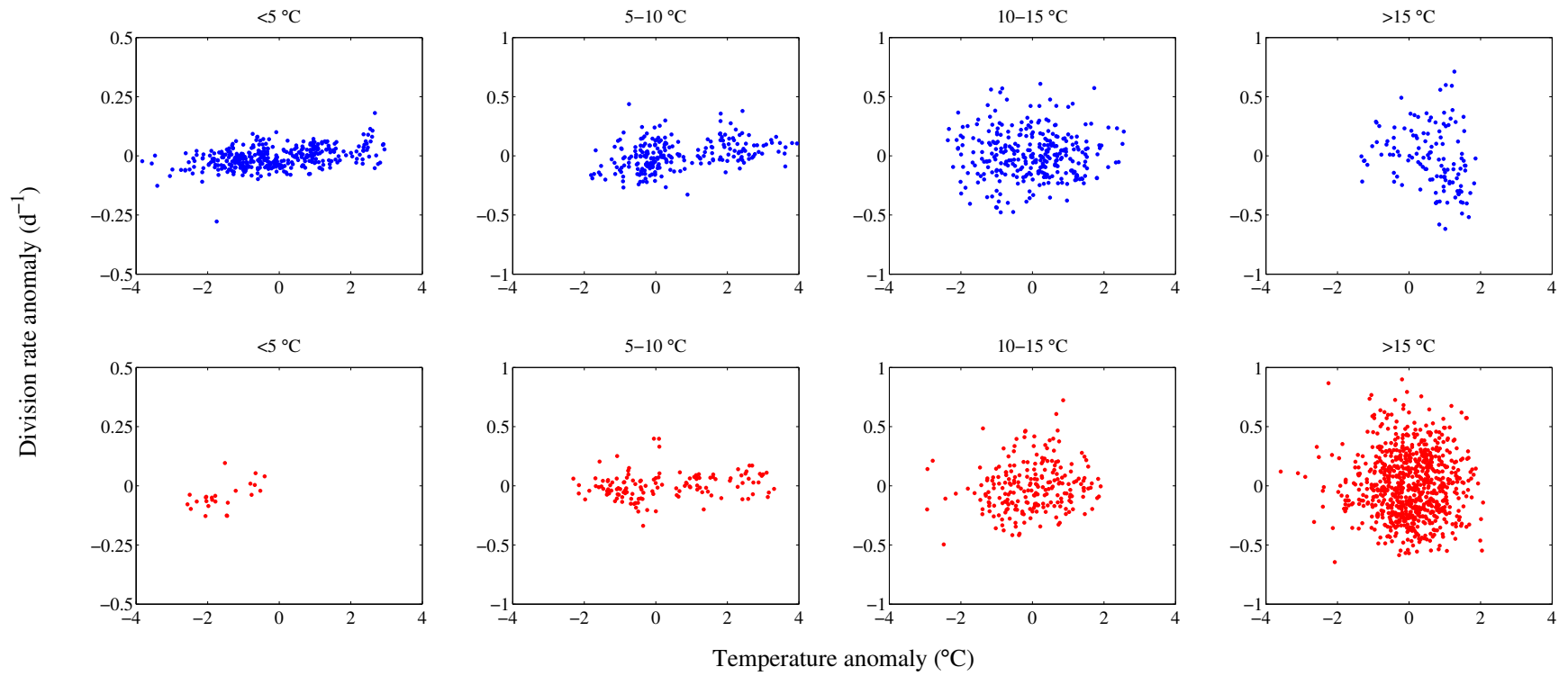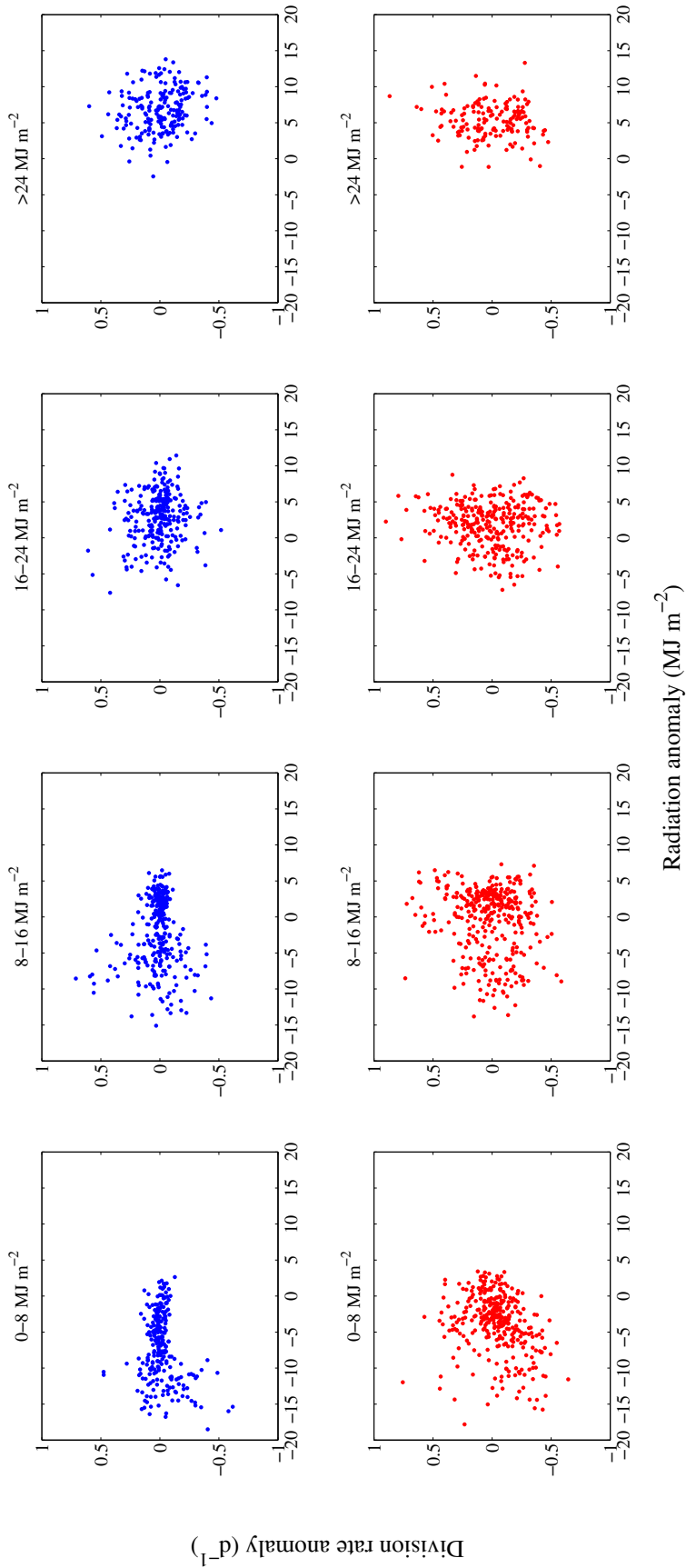
Figure A-19: The top row shows the relationships between anomalies of division rate and anomalies of temperature for different temperature intervals for the first half of the year (January - June) (indicated as blue points), while the bottom row shows these relationships for the second half of the year (July-December) (indicated as red points).

Figure A-20: The top row shows the relationships between anomalies of *Synechococcus* cell concentration and anomalies of solar radiation for different radiation intervals for the first half of the year (January - June) (indicated as blue points), while the bottom row shows these relationships for the second half of the year (July-December) (indicated as red points).

Figure A-21: The top row shows the relationships between anomalies of division rate and anomalies of solar radiation for different radiation intervals for the first half of the year (January - June) (indicated as blue points), while the bottom row shows these relationships for the second half of the year (July-December) (indicated as red points).
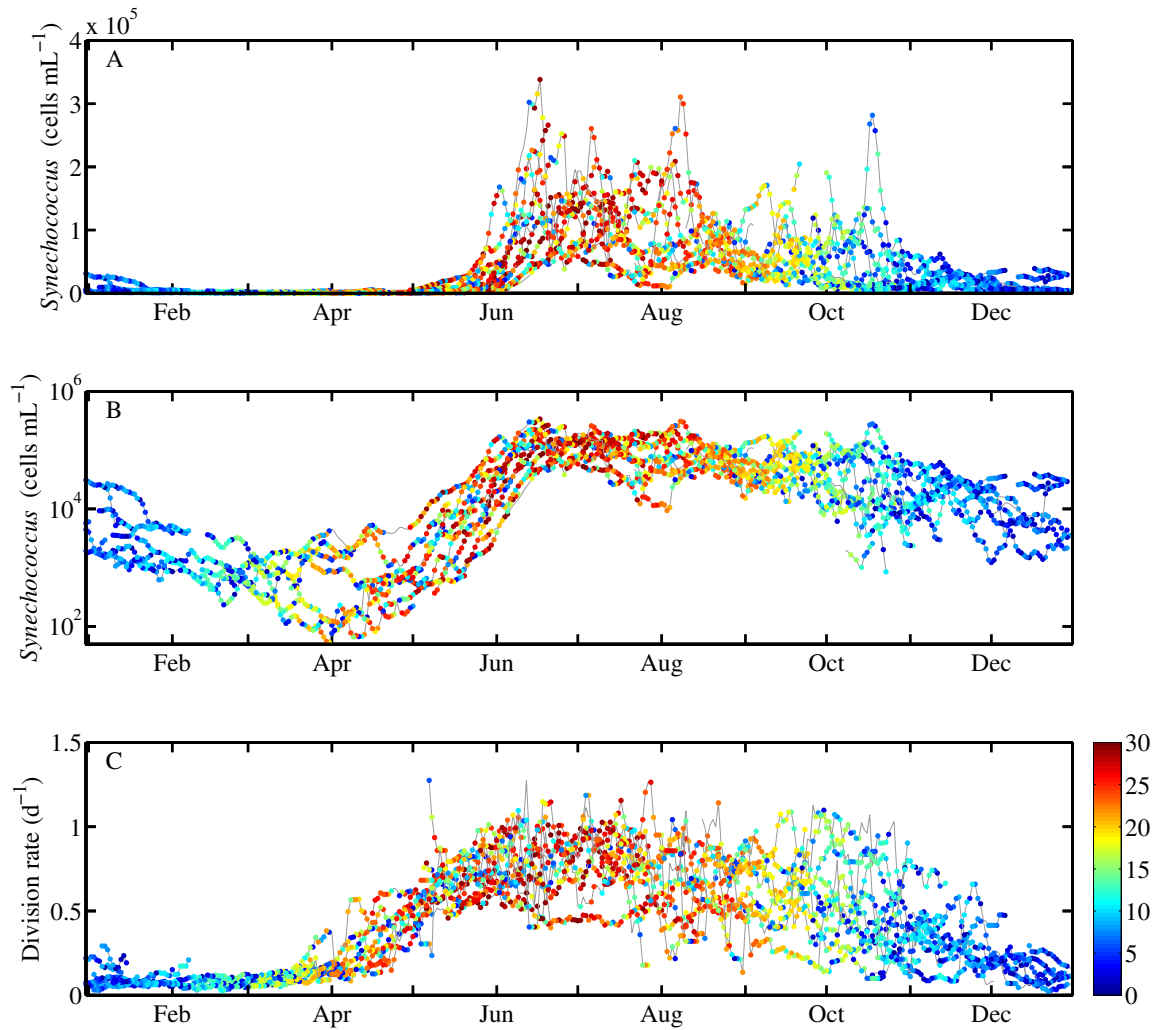
171
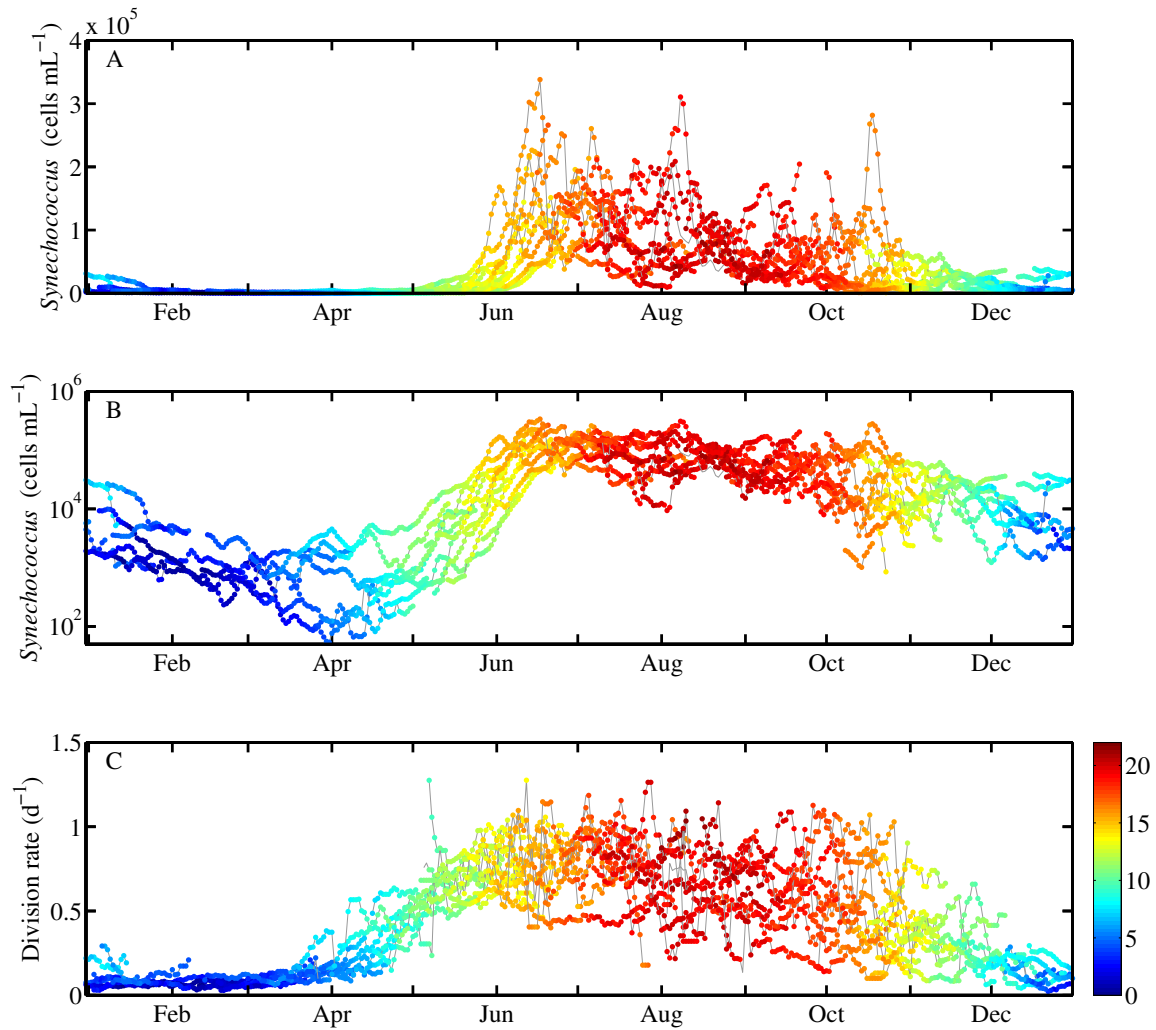
Figure A-22: A. Overlay of daily averages of *Synechococcus* cell concentration for each year plotted against year day. Color indicates temperature for that day. B. Overlay of daily averages of *Synechococcus* cell concentration for each year plotted against year day, but represented on a log scale. C. Overlay of daily division rate estimates for each year plotted against year day.

Figure A-23: A. Overlay of daily averages of *Synechococcus* cell concentration for each year plotted against year day. Color indicates daily solar radiation. B. Overlay of daily averages of *Synechococcus* cell concentration for each year plotted against year day, but represented on a log scale. C. Overlay of daily division rate estimates for each year plotted against year day.
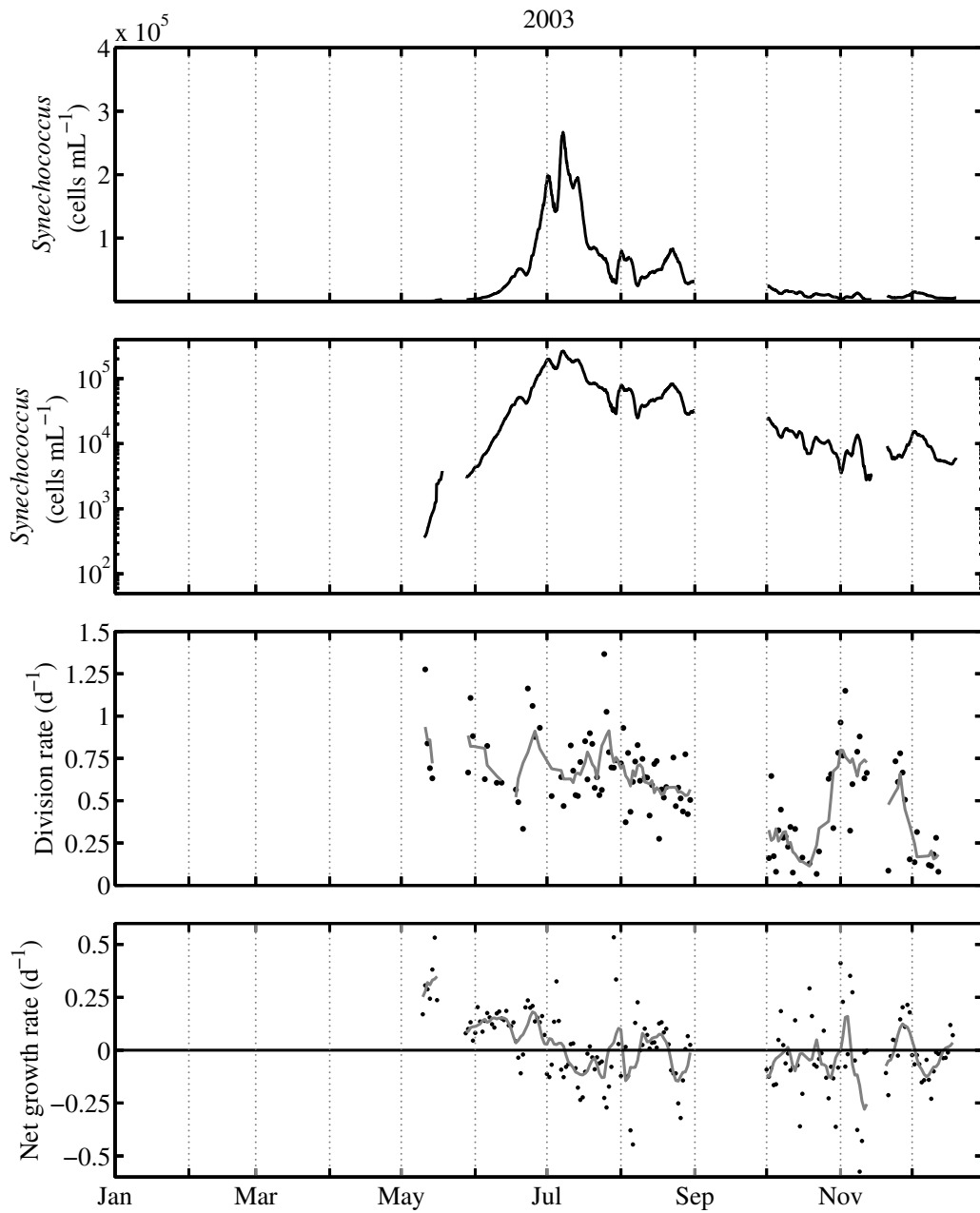
Figure A-24: A. Smoothed *Synechococcus* cell concentration for 2003 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running means.
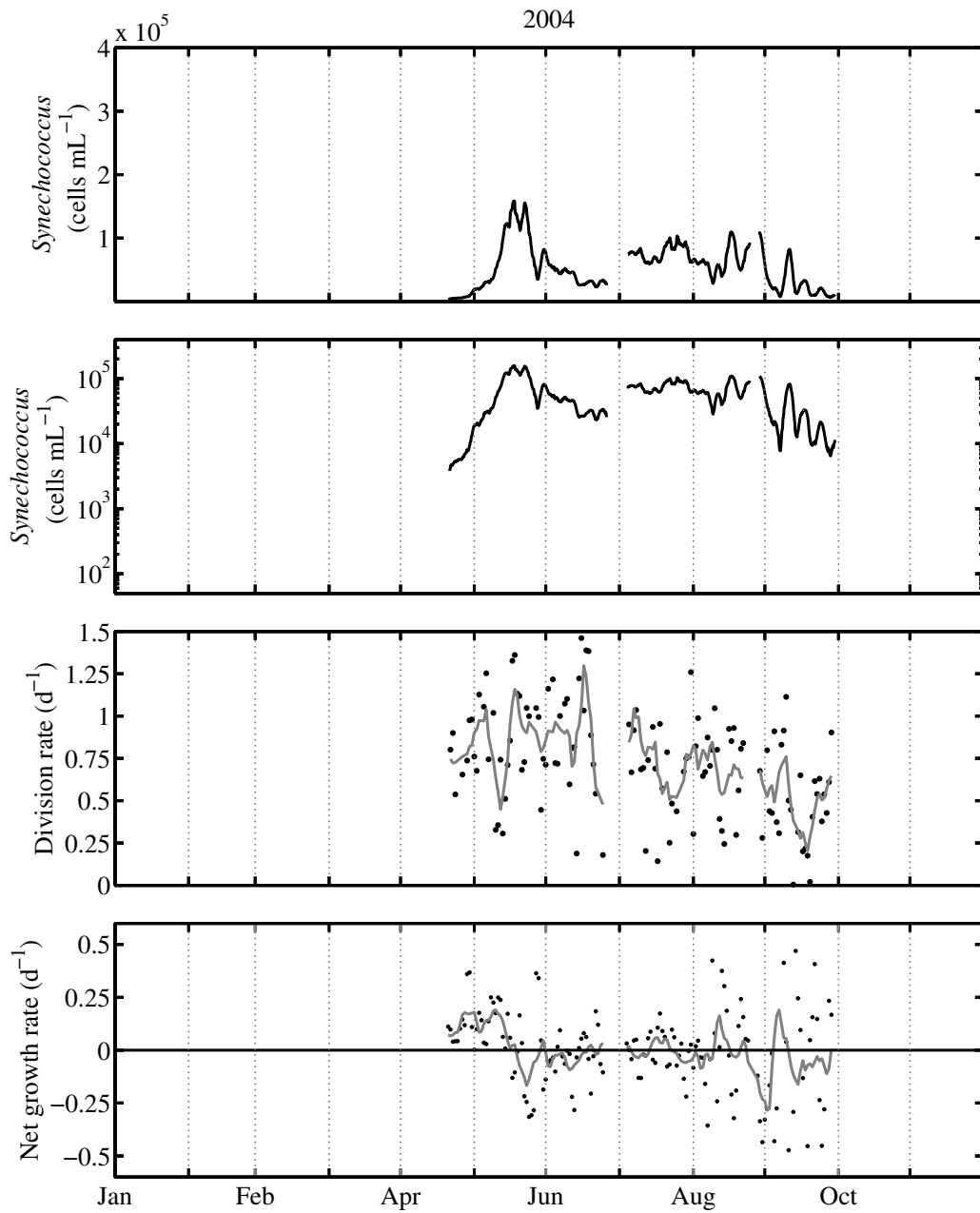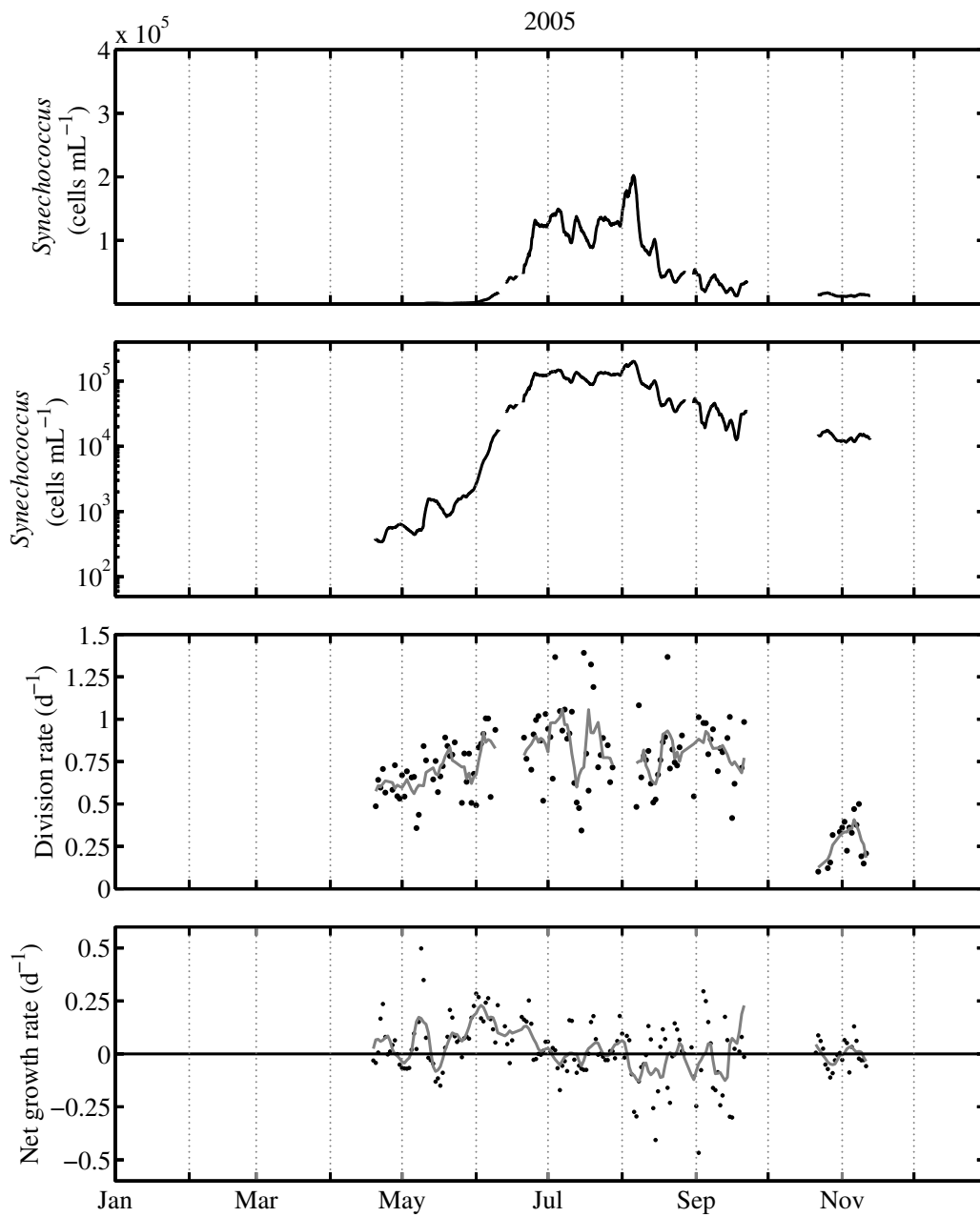
Figure A-25: A. Smoothed *Synechococcus* cell concentration for 2004 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running means.

Figure A-26: A. Smoothed *Synechococcus* cell concentration for 2005 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running mean.
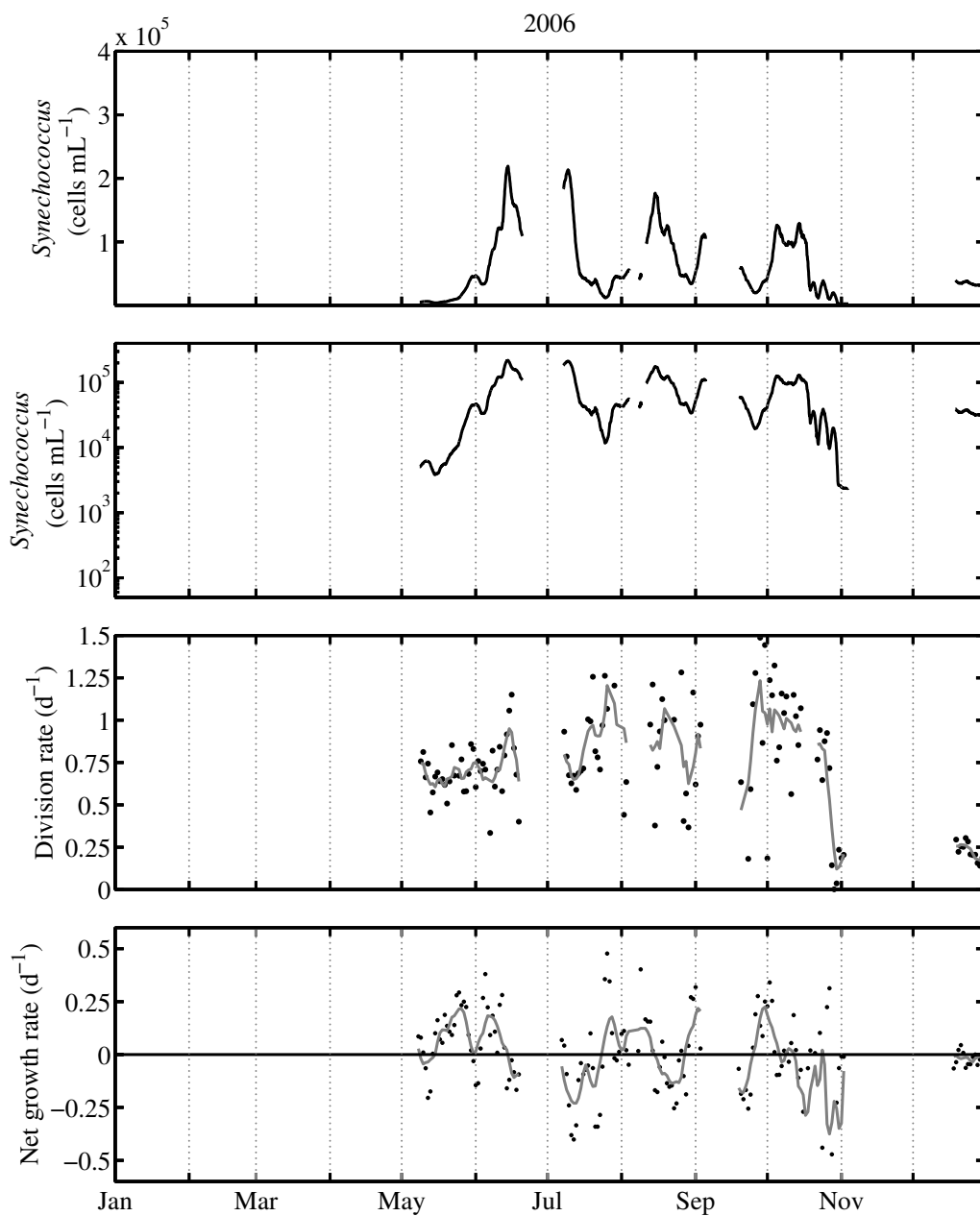
Figure A-27: A. Smoothed *Synechococcus* cell concentration for 2006 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running mean.
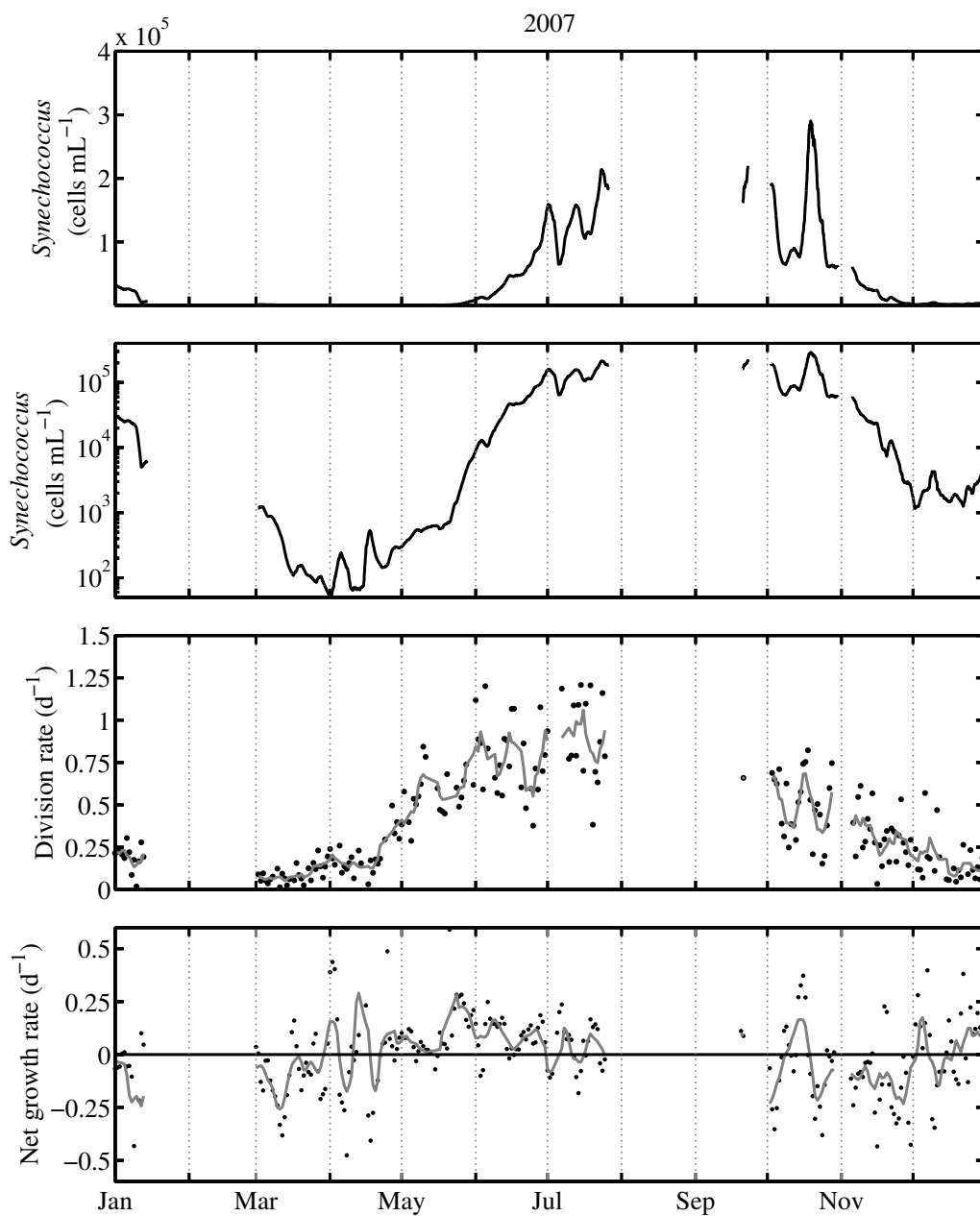
Figure A-28: A. Smoothed *Synechococcus* cell concentration for 2007 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running mean.
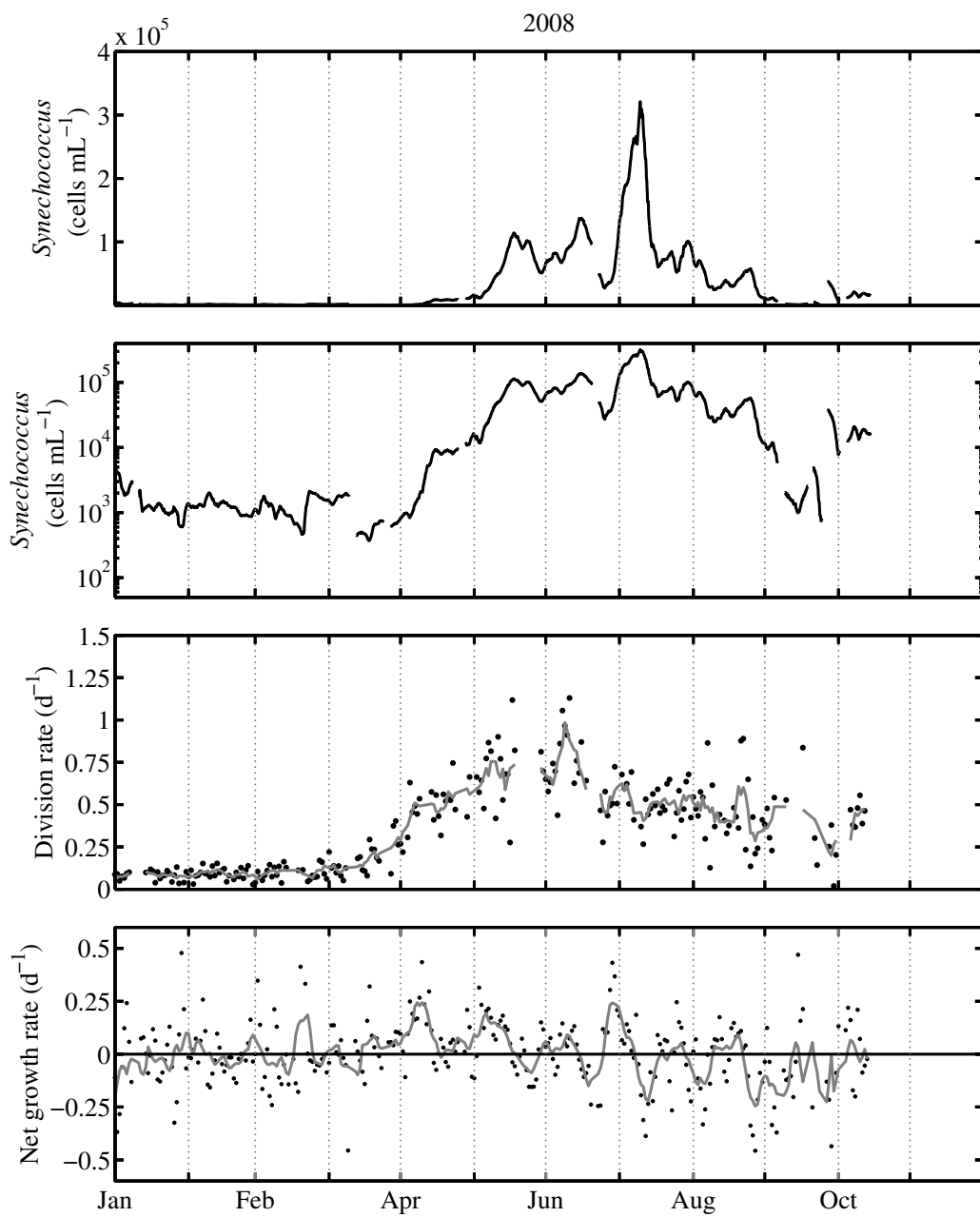
Figure A-29: A. Smoothed *Synechococcus* cell concentration for 2008 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running mean.
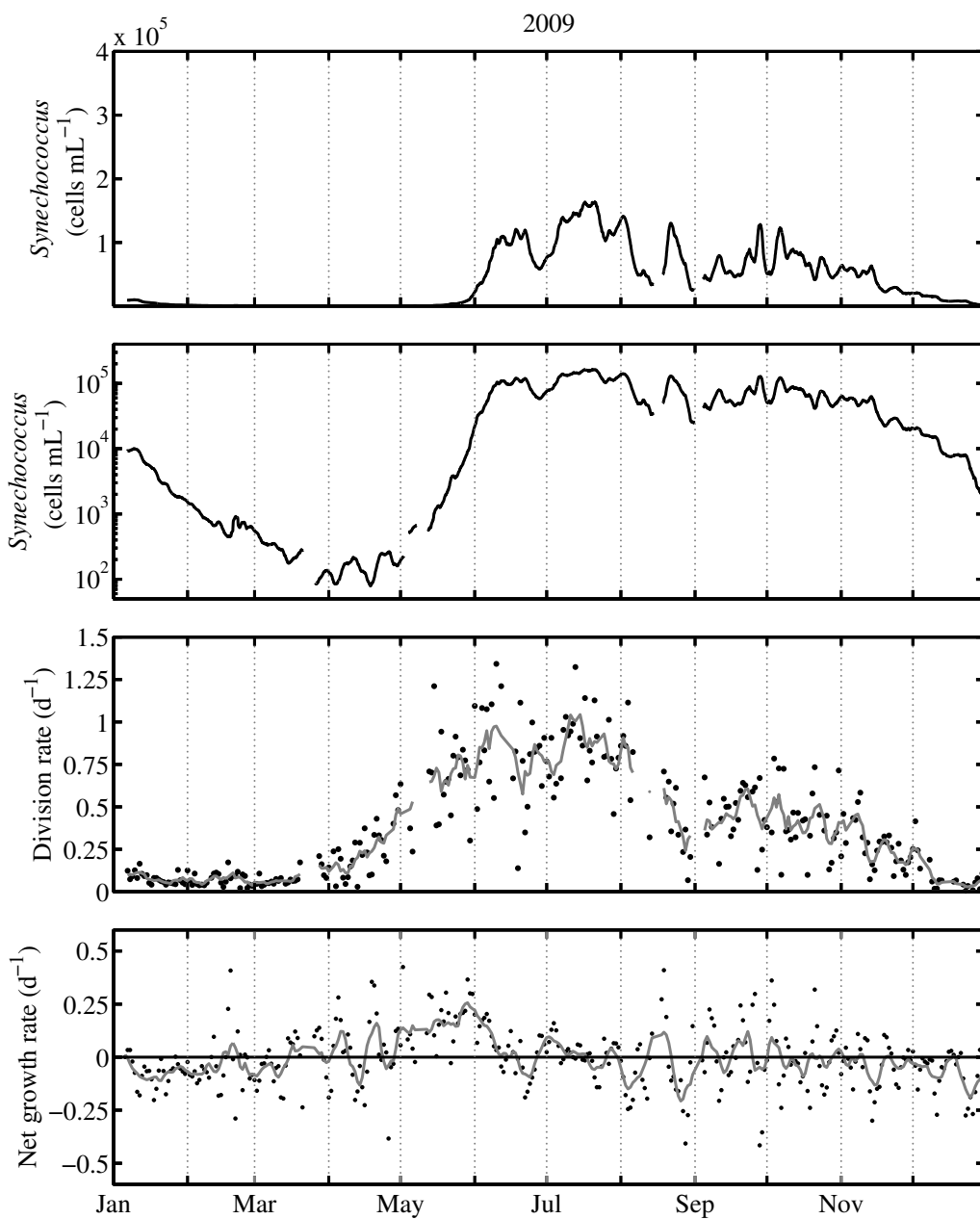
Figure A-30: A. Smoothed *Synechococcus* cell concentration for 2009 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running mean.

Figure A-31: A. Smoothed *Synechococcus* cell concentration for 2010 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running mean.
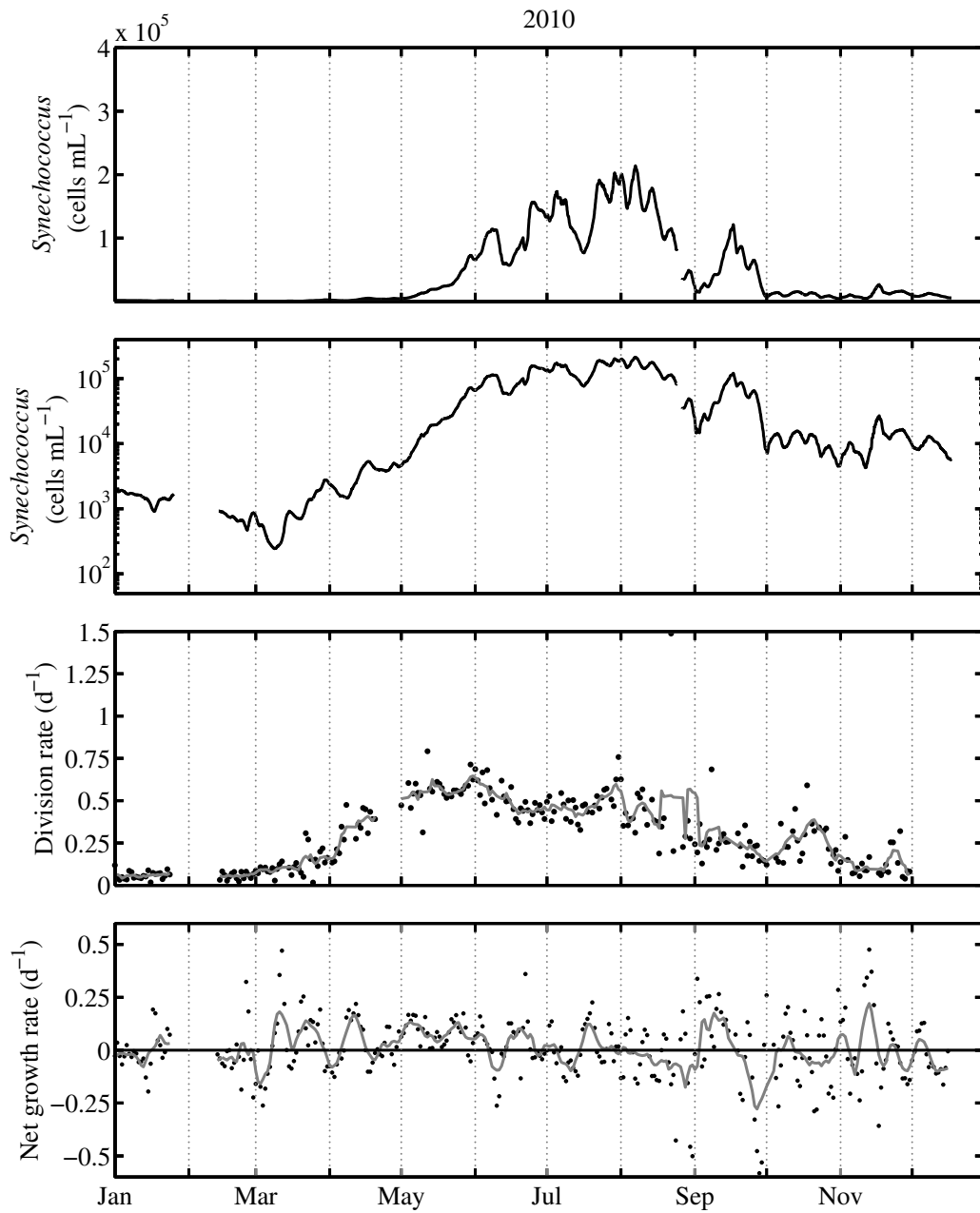
Figure A-32: A. Smoothed *Synechococcus* cell concentration for 2011 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running mean.

Figure A-33: A. Smoothed *Synechococcus* cell concentration for 2012 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running mean.
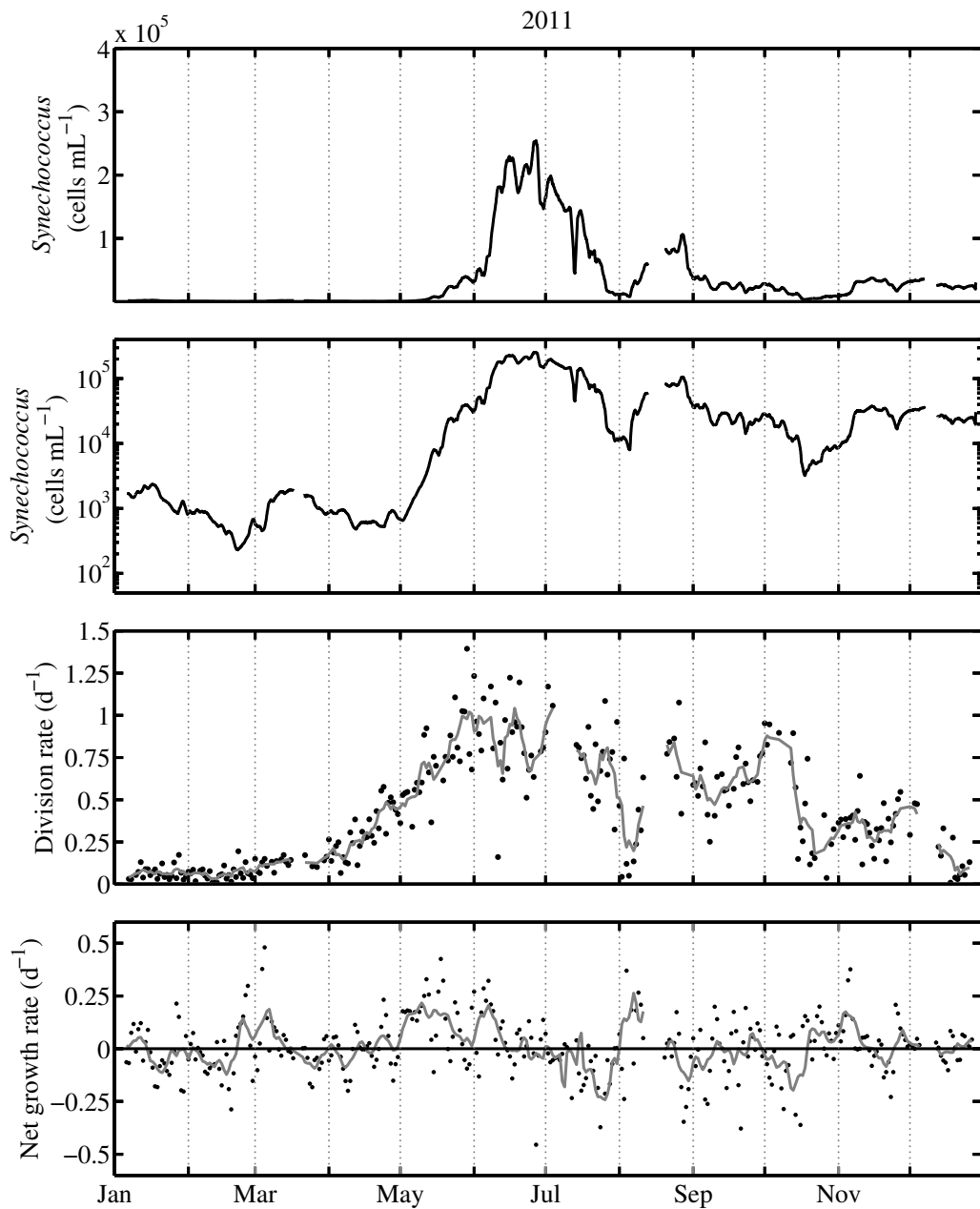
Figure A-34: A. Smoothed *Synechococcus* cell concentration for 2013 (48-hr running mean). B. Smoothed *Synechococcus* cell concentration as presented in panel A but shown on a log scale. C. Daily division rate estimates. D. Daily net growth rate. Gray lines are 7-day running mean.
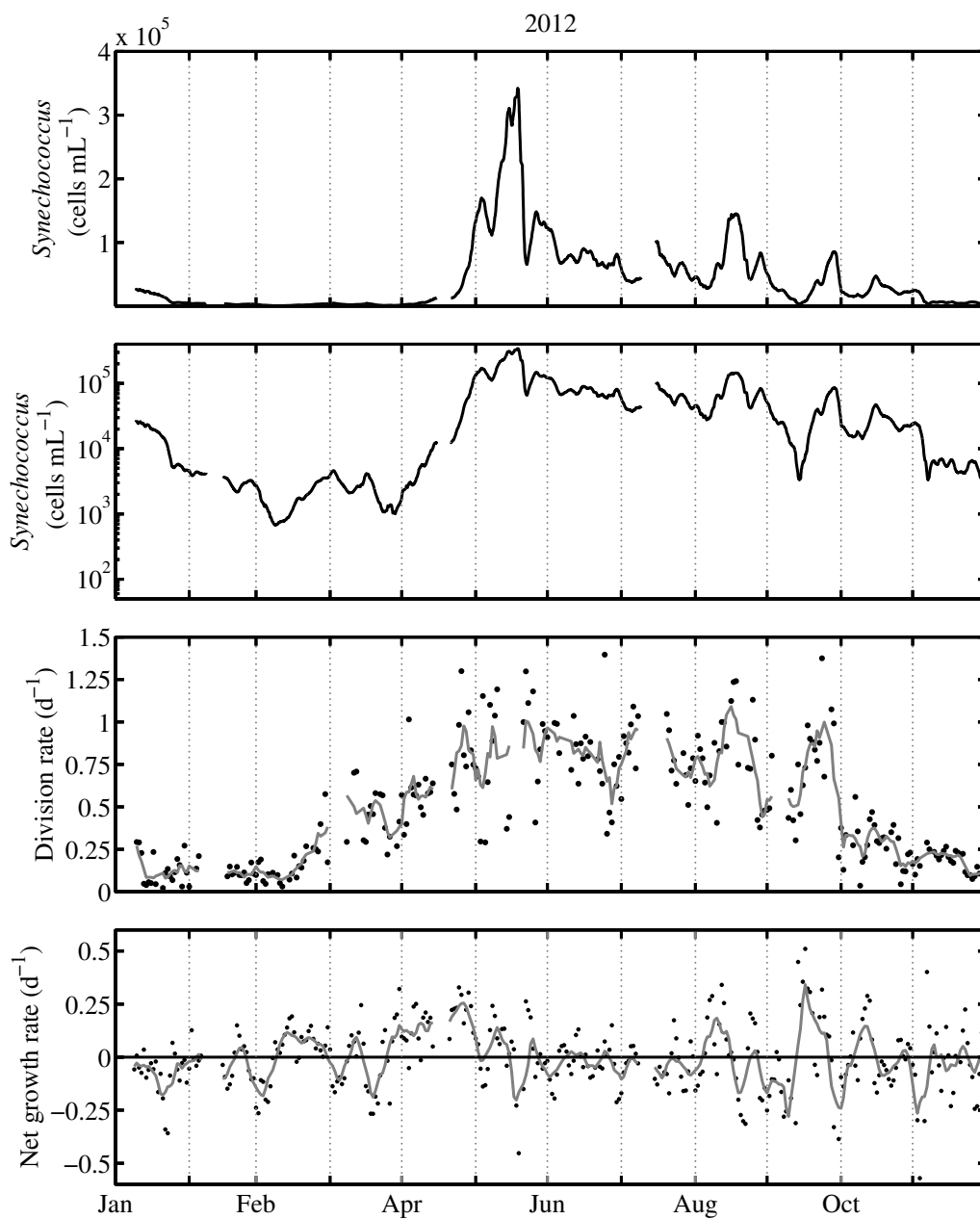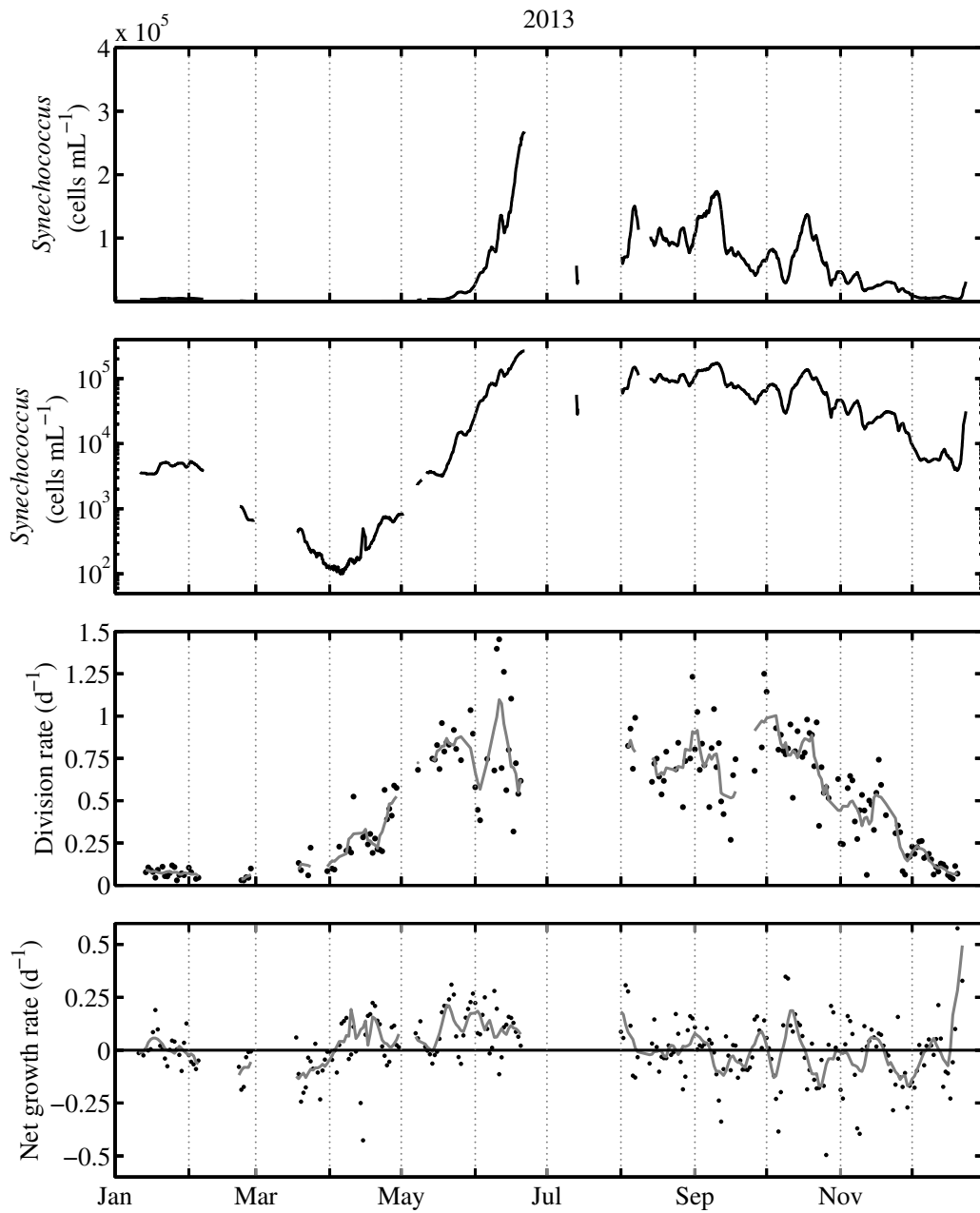
Appendix for Chapter 4

Table A.2: Table of strains isolated from MVCO for which *ntcA* gene was sequenced. Strains are organized by clade designation and any duplicate strain (i.e., strain with an identical *ntcA* sequence) are listed with their dates of isolation in columns 4 and 5. Pigment type is indicated as either phycoerythrin (PE) or phycocyanin (PC). Asterisk indicates cultures that also had 16S rRNA gene sequenced (see Chapter 5).

Clade	Strain Name	Date	Duplicate Strains	Major Date	Phycobiliprotein
IC	MV0801*	11/22/11			PE
IC			MV1001	02/08/12	PE
IC	MV1118*	03/06/12			PE
IC			MV1202B	04/20/12	PE
IC			MV1202X*	04/20/12	PE
IC			MV1205	04/20/12	PE
IC			MV1217	04/20/12	PE
IC			MV1408*	07/10/12	PE
IC	MV0917*	01/09/12			PE
IC			MV0902	01/09/12	PE
IC			MV1220	04/20/12	PE
IC			MV1310	06/15/12	PE
IC			MV1305	06/15/12	PE
IC			MV1417	07/10/12	PE
IC			MV1416A	07/10/12	PE
IC	MV0913	01/09/12			PE
IC			MV0914A	01/09/12	PE
IC			MV0909	01/09/12	PE
IC			MV1309A	06/15/12	PE
IC			MV1302*	06/15/12	PE
IC			MV1418	07/10/12	PE
IC	MV0906	01/09/12			PE
IC	MV1106	03/06/12			PE
IC	MV1109	03/06/12			PE
IC	MV1308*	06/15/12			PE
IC	MV1309U	06/15/12			PE
IC	MV1416W	07/10/12			PE
IC	MV1212	04/20/12			PE
IC	MV0919*	01/09/12			PE
IE	MV0403	09/12/11			PE
IE			MV0415E*	09/12/11	PE
IE			MV0416E	09/12/11	PE
IE			MV0416D	09/12/11	PE

IE			MV0411	09/12/11	PE
IE			MV0415D	09/12/11	PE
IE			MV0615	10/06/11	PE
IE			MV0804	11/22/11	PE
IE			MV0802X*	11/22/11	PE
IE			MV0802Y	11/22/11	PE
IE			MV0918	01/09/12	PE
IE			MV0914B	01/09/12	PE
IE			MV1605A	09/14/12	PE
IE			MV1609	09/14/12	PE
IE			MV1706	10/08/12	PE
IE			MV1705*	10/08/12	PE
IE			MV1718	10/08/12	PE
IE			MV1701	10/08/12	PE
IE	MV1307*	06/15/12			PE
IE			MV1716	10/08/12	PE
IE	MV1717	10/08/12			PE
IE	MV0210*	07/27/11			PE
IE			MV0203RD	07/27/11	PE
IE			MV0203RF	07/27/11	PE
IE			MV1311	06/15/12	PE
IE			MV1318	06/15/12	PE
IE			MV1320*	06/15/12	PE
IE	MV0901X	01/09/12			PE
IE	MV1112*	03/06/12			PE
IE	MV0606D	10/06/11			PE
IE	MV0908*	01/09/12			PE
IE	MV1306	06/15/12			PE
IE	MV1610	09/14/12			PE
IE	MV1301	06/15/12			PE
IE	MV0326	08/20/11			PE
II	MV0519B*	09/27/11			PE
II			MV1712	10/08/12	PE
II	MV1720	10/08/12			PE
II	MV1604*	09/14/12			PE
III	MV1520	08/20/12			PE
III	MV1602	09/14/12			PE
III	MV1605B*	09/14/12			PE
VI	MV0417*	09/12/11			PE
VI			MV0420	09/12/11	PE
VI			MV1420	07/10/12	PE
VI			MV1519E	08/20/12	PE
VI			MV1519B	08/20/12	PE
VI			MV1710	10/08/12	PE
VI	MV1515	08/20/12			PE
VI	MV0507	09/27/11			PE
VI	MV1612	09/14/12			PE
VI	MV0513	09/27/11			PE
VI			MV0608E	10/06/11	PE
VI			MV0608I	10/06/11	PE

VI			MV0619	10/06/11	PE
VI			MV0613A	10/06/11	PE
VI			MV0613C	10/06/11	PE
VI			MV1614B	09/14/12	PE
VI			MV1611	09/14/12	PE
VI			MV1618C*	09/14/12	PE
VI			MV1618E	09/14/12	PE
VI			MV1613	09/14/12	PE
VI	MV0501D	09/27/11			PE
VI			MV0510B	09/27/11	PE
VI			MV0510BF	09/27/11	PE
VI			MV0612	10/06/11	PE
VI			MV1619	09/14/12	PE
VI			MV1615	09/14/12	PE
VI			MV1616	09/14/12	PE
VI	MV1614C	09/14/12			PE
VI	MV1617	09/14/12			PE
VII	MV0516	09/27/11			PE
VII			MV0609B*	10/06/11	PE
VII			MV0609I	10/06/11	PE
VII			MV0606C	10/06/11	PE
VII			MV0611	10/06/11	PE
VII			MV0711	10/31/11	PE
VII	MV1607A*	09/14/12			PE
VII			MV1607B	09/14/12	PE
VIII	MV1512	08/20/12			PC
VIII	MV1620*	09/14/12			PC
CB4	MV1312	06/15/12			PC
CB4	MV1213	04/20/12			PC
CB4	MV0409*	09/12/11			PC
CB4			MV0506	09/27/11	PC
CB4	MV0504	09/27/11			PC
CB4	MV0510BG	09/27/11			PC
CB5	MV0414	09/12/11			PE
CB5	MV0605E*	10/06/11			PE
CB5	MV0601D	10/06/11			PE
5.2MV1	MV0305	08/20/11			PC
5.2MV1	MV0143	07/06/11			PC
5.2MV1			MV0706	10/31/11	PC
5.2MV1			MV0709	10/31/11	PC
5.2MV1			MV0816	11/22/11	PC
5.2MV1			MV0820	11/22/11	PC
5.2MV1			MV1010	02/08/12	PC
5.2MV1			MV1006*	02/08/12	PC
5.2MV1	MV1117	03/06/12			PC
5.2MV1	MV0216	07/27/11			PC
5.2MV1	MV0203B	07/27/11			PC
5.2MV1	MV0704*	10/31/11			PC
5.2MV1	MV0715	10/31/11			PC
5.2MV1	MV0501F	09/27/11			PC

5.2MV2	MV1218D *	04/20/12				PE
5.2MV2			MV1218C	04/20/12		PE
5.2MV3	MV1219	04/20/12				PC
5.2MV3	MV1215	04/20/12				PC
5.2MV3	MV0418	09/12/11				PC
5.2MV3			MV0402	09/12/11		PC
5.2MV3	MV0910*	01/09/12				PC
5.2MV3	MV0116	07/06/11				PC
5.3I	MV1715*	10/08/12				PE
5.3I	MV0610	10/06/11				PE
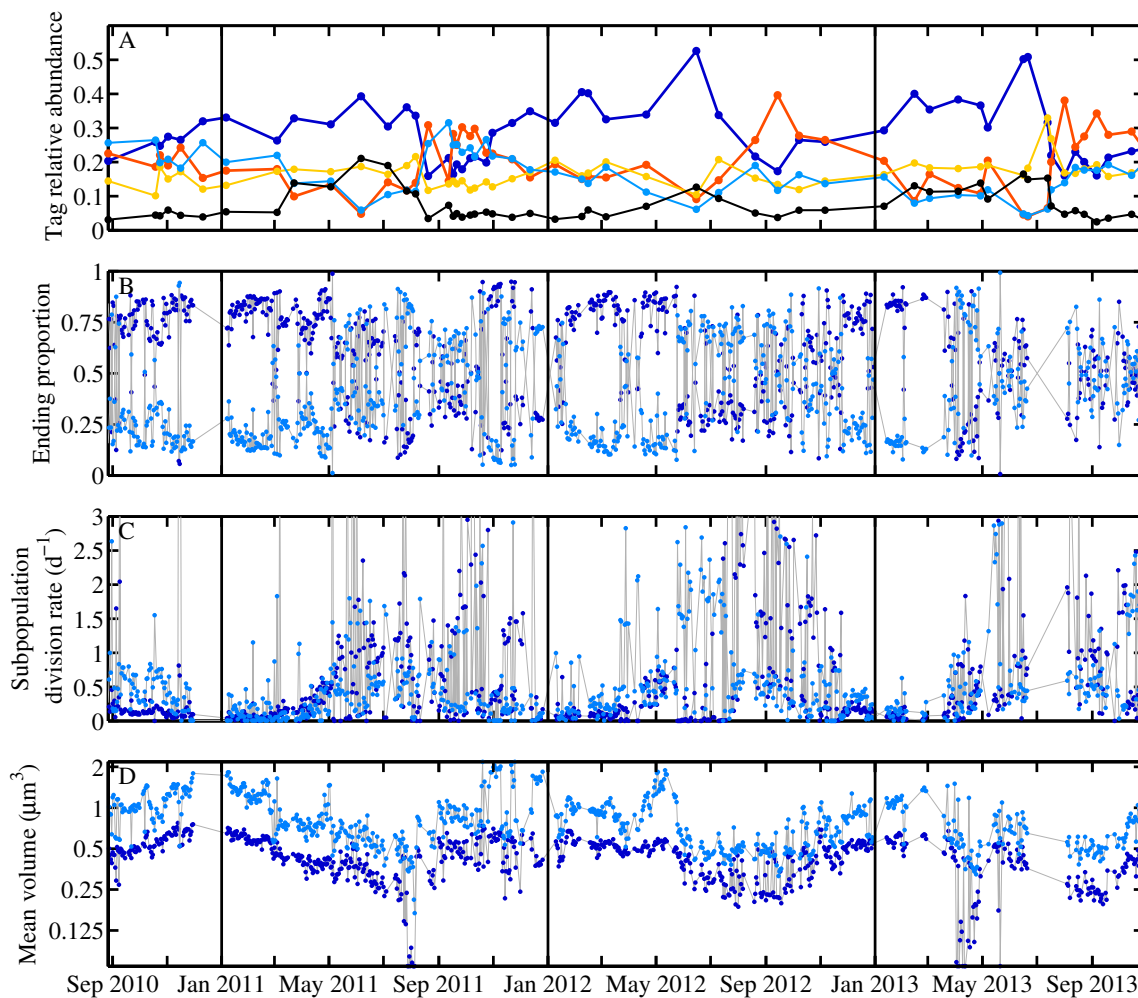
Figure A-35: A. Relative abundances of the top 5 oligotypes. Colors are same as in Fig. 5-3 (O1-I in dark blue, O2-M in orange, O3-M in gold, O4-I in light blue, and O5-I* in black) B. Ending proportion of each subpopulation in the model. Subpopulations are based on starting cell volume. Small volume populations are denoted by dark blue, while larger volume populations are denoted by light blue. C. Subpopulation division rates. D. Mean cell volumes fit by the model for each subpopulation for start of each day.

Bibliography

[1] Nona SR Agawin, Carlos M Duarte, and Susana Agusti. Growth and abundance of *Synechococcus* sp. in a Mediterranean Bay: seasonality and relationship with temperature. *Marine Ecology Progress Series*, 170:45–53, 1998.

[2] Nona SR Agawin, Carlos M Duarte, Susana Agustí, et al. Response of Mediterranean *Synechococcus* growth and loss rates to experimental nutrient inputs. *Marine Ecology Progress Series*, 206:97–106, 2000.

[3] Nathan A. Ahlgren and Gabrielle Rocap. Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Applied and Environmental Microbiology*, 72(11):7193–7204, November 2006.

[4] Nathan A Ahlgren and Gabrielle Rocap. 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*, 3:213, 2012.

[5] Nathan A Ahlgren, Gabrielle Rocap, and Sallie W Chisholm. Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environmental Microbiology*, 8(3):441–454, 2006.

[6] Jude K Apple, Suzanne L Strom, Brian Palenik, and Bianca Brahamsha. Variability in protist grazing and growth on different marine *Synechococcus* isolates. *Applied and Environmental Microbiology*, 77(9):3074–3084, 2011.

[7] Adelchi Azzalini. Statistical Inference Based on the likelihood. In *Monographs on Statistics and Applied Probability*, volume 68. Chapman and Hall, 1996.

[8] A.C. Baudoux, M.J.W. Veldhuis, A.A.M. Noordeloos, G. van Noort, and C.P.D. Brussaard. Estimates of virus- vs. grazing induced mortality of picophytoplankton in the North Sea during summer. *Aquatic Microbial Ecology*, 52:69–82, July 2008.

[9] Anne-Claire Baudoux, Marcel JW Veldhuis, Harry J Witte, and Corina PD Brussaard. Viruses as mortality agents of picophytoplankton in the deep chlorophyll maximum layer during IRONAGES III. *Limnology and Oceanography*, 52(6):2519–2529, 2007.

[10] Brian J. Binder and Sallie W. Chisholm. Cell cycle regulation in marine *Synechococcus* sp. strains. *Applied and Environmental Microbiology*, 61(2):708–717, February 1995.

[11] Brian J. Binder, Sallie W. Chisholm, Robert J. Olson, Sheila L. Frankel, and Alexandra Z. Worden. Dynamics of picophytoplankton, ultraphytoplankton and bacteria in the central equatorial Pacific. *Deep-Sea Research II*, 43(4-6):907–931, 1996.

[12] Haiyuan Cai, Kui Wang, Sijun Huang, Nianzhi Jiao, and Feng Chen. Distinct patterns of picocyanobacterial communities in winter and summer in the Chesapeake Bay. *Applied and Environmental Microbiology*, 76(9):2955–2960, 2010.

[13] Lisa Campbell and Edward J. Carpenter. Estimating the grazing pressure of heterotrophic nanoplankton on *Synechococcus* spp. using the sea water dilution and selective inhibitor techniques. *Marine Ecology Progress Series*, 33:121–129, 1986.

[14] Lisa Campbell, Robert J. Olson, Heidi M. Sosik, Ann Abraham, Darren W. Henrichs, Cammie J. Hyatt, and Edward J. Buskey. First harmful *Dinophysis* (Dinophyceae, Dinophysiales) bloom in the U.S. is revealed by automated imaging flow cytometry. *Journal of Phycology*, 46:66–75, 2010.

[15] David A Caron, Emily R Peele, Ee Lin Lim, Mark R Dennett, et al. Picoplankton and nanoplankton and their trophic coupling in surface waters of the Sargasso Sea south of Bermuda. *Limnology and Oceanography*, 44(2):259–272, 1999.

[16] Hal Caswell. *Matrix Population Models*. Sinauer Associates, 2nd edition, 2001.

[17] Feng Chen, Kui Wang, Jinjun Kan, Marcelino T Suzuki, and K Eric Wommack. Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Applied and Environmental Microbiology*, 72(3):2239–2243, 2006.

[18] Dong Han Choi and Jae Hoon Noh. Phylogenetic diversity of *Synechococcus* strains isolated from the East China Sea and the East Sea. *FEMS Microbiology Ecology*, 69(3):439–448, 2009.

[19] Joon W Choi and Francesc Peters. Effects of temperature on two psychrophilic ecotypes of a heterotrophic nanoflagellate, paraphysomonas imperforata. *Applied and Environmental Microbiology*, 58(2):593–599, 1992.

[20] U Christaki, C Courties, H Karayanni, A Giannakourou, C Maravelias, K Ar Kormas, and P Lebaron. Dynamic characteristics of *Prochlorococcus* and*Synechococcus* consumption by bacterivorous nanoflagellates. *Microbial Ecology*, 43(3):341–352, 2002.

[21] Urania Christaki, Stéphan Jacquet, John R Dolan, Daniel Vaulot, and Ferei-doun Rassoulzadegan. Growth and grazing on *Prochlorococcus and* Synechococ-cus by two marine ciliates. *Limnology and Oceanography*, 44(1):52–61, 1999.

[22] James H Churchill, Peter C Cornillon, and George W Milkowski. A cyclonic eddy and shelf-slope water exchange associated with a Gulf Stream warm-core ring. *Journal of Geophysical Research*, 91(C8):9615–9623, 1986.

[23] G. Dall'Olmo, E. Boss, Michael J. Behrenfeld, T. K. Westberry, C. Courties, L. Prieur, M. Pujo-Pay, N. Hardman-Mountford, and T. Moutin. Inferring phytoplankton carbon and eco-physiological rates from diel cycles of spectral particulate beam-attentuation coefficient. *Biogeosciences*, 8:3423–3440, 2011.

[24] Alexis Dufresne, Martin Ostrowski, David J Scanlan, Laurence Garczarek, So-phie Mazard, Brian P Palenik, Ian T Paulsen, N Tandeau de Marsac, Patrick Wincker, Carole Dossat, Steve Ferriera, Justin Johnson, Anton F Post, Wolf-gang R Hess, and Frederic Partensky. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biology*, 9(5):R90, 2008.

[25] Michelle D. DuRand. *Phytoplankton growth and diel variations in beam atten-uation through individual cell analysis*. PhD thesis, Massachusetts Institute of Technology, Woods Hole Oceanographic Institution, 1995.

[26] Paul G. Falkowski and John A. Raven. *Aquatic Photosynthesis*. Princeton University Press, 2nd edition, 2007.

[27] Christopher B. Field, Michael J. Behrenfeld, James T. Randerson, and Paul Falkowski. Primary production of the biosphere: Intergrating terrestrial and oceanic components. *Science*, 281:237–240, July 1998.

[28] Nicholas J Fuller, Dominique Marie, Frédéric Partensky, Daniel Vaulot, Anton F Post, and David J Scanlan. Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Applied and Environmental Microbiol-ogy*, 69(5):2430–2443, 2003.

[29] Charles L. Gallegos. Microzooplankton grazing in phytoplankton in the Rhode River, Maryland: nonlinear feeding kinetics. *Marine Ecology Progress Series*, 57:23–33, 1989.

[30] Glen Gawarkiewicz, Frank Bahr, Robert C Beardsley, and Kenneth H Brink. Interaction of a slope eddy with the shelfbreak front in the Middle Atlantic Bight. *Journal of Physical Oceanography*, 31(9):2783–2796, 2001.

[31] Glen Gawarkiewicz, Kenneth H Brink, Frank Bahr, Robert C Beardsley, Michael Caruso, James F Lynch, and Ching-Sang Chiu. A large-amplitude meander of the shelfbreak front during summer south of New England: Ob-servations from the shelfbreak PRIMER experiment. *Journal of Geophysical Research*, 109(C03006), 2004.

[32] Juan M. Gonzalez, Evelyn B. Sherr, and Barry F. Sherr. Size-selective grazing on bacteria by natural assemblaages of estuarine flagellates and ciliates. *Applied and Environmental Microbiology*, 56(3):583–589, 1990.

[33] Tom A Hall. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41:95–98, 1999.

[34] Thomas Haverkamp, Silvia G Acinas, Marije Doeleman, Maayke Stomp, Jef Huisman, and Lucas J Stal. Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environmental Microbiology*, 10(1):174–188, 2008.

[35] Thomas HA Haverkamp, Daphne Schouten, Marije Doeleman, Ute Wollenzien, Jef Huisman, and Lucas J Stal. Colorful microdiversity of *Synechococcus* strains (picocyanobacteria) isolated from the Baltic Sea. *The ISME Journal*, 3(4):397–408, 2009.

[36] David V Hinkley. Inference about the intersection in two-phase regression. *Biometrika*, 56(3):495–504, 1969.

[37] Sijun Huang, Steven W Wilhelm, H Rodger Harvey, Karen Taylor, Nianzhi Jiao, and Feng Chen. Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *The ISME Journal*, 6(2):285–297, 2012.

[38] Kristen R Hunter-Cevera, Michael G Neubert, Andrew R Solow, Robert J Olson, Alexi Shalapyonok, and Heidi M Sosik. Diel size distributions reveal seasonal growth dynamics of a coastal phytoplankter. *Proceedings of the National Academy of Sciences*, 111(27):9852–9857, 2014.

[39] Susan M Huse, Les Dethlefsen, Julie A Huber, David Mark Welch, David A Relman, and Mitchell L Sogin. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS genetics*, 4(11):e1000255, 2008.

[40] Susan M Huse, David B Mark Welch, Andy Voorhis, Anna Shipunova, Hilary G Morrison, A Murat Eren, and Mitchell L Sogin. VAMPS: a website for visualization and analysis of microbial population structures. *BMC bioinformatics*, 15(1):41, 2014.

[41] G Evelyn Hutchinson. The paradox of the plankton. *American Naturalist*, pages 137–145, 1961.

[42] Stephan Jacquet, Frederic Partensky, Jean-Francois Lennon, and Daniel Vaulot. Diel patterns of growth and division in marine picoplankton in culture. *Journal of Phycology*, 37:357–369, June 2001.

[43] Ludwig Jardillier, Mikhail V Zubkov, John Pearman, and David J Scanlan. Significant CO_2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *The ISME Journal*, 4(9):1180–1192, 2010.

[44] Hae Jin Jeong, Jae Yeon Park, Jae Hoon Nho, Myung Ok Park, Jeong Hyun Ha, Kyeong Ah Seong, Chang Jeng, Chi Nam Seong, Kwang Ya Lee, and Won Ho Yih. Feeding by red-tide dinoflagellates on the cyanobacterium *Synechococcus*. *Aquatic Microbial Ecology*, 41(2):131–143, 2005.

[45] Norman Lloyd Johnson and Samuel Kotz. *Distributions in statistics: continuous multivariate distributions*. Wiley, 1972.

[46] Paul W Johnson and JM Sieburth. Chroococcoid cyanobacteria in the sea: a ubiquitous and diverse phototrophic biomass. *Limnology and Oceanography*, 24(5):928–935, 1979.

[47] Zackary I Johnson, Erik R Zinser, Allison Coe, Nathan P McNulty, E Malcolm S Woodward, and Sallie W Chisholm. Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science*, 311(5768):1737–1740, 2006.

[48] Terrence M Joyce, James KB Bishop, and Otis B Brown. Observations of offshore shelf-water transport induced by a warm-core ring. *Deep Sea Research Part A. Oceanographic Research Papers*, 39:S97–S113, 1992.

[49] Klaus Jurgens and Carsten Matz. Predation as a shaping force for the phenotypic and genotypic composition of planktonic bacteria. *Antonie van Leeuwenhoek*, 81:413–434, 2002.

[50] Anthony R Kirincich, Steven J Lentz, J Thomas Farrar, and Neil K Ganju. The spatial structure of tidal and mean circulation over the inner shelf south of Martha's Vineyard, Massachusetts. *Journal of Physical Oceanography*, 43(9):1940–1958, 2013.

[51] Mark Kot. *Elements of Mathematical Ecology*. Cambridge University Press, 2001.

[52] Michael R Landry and Albert Calbet. Microzooplankton production in the oceans. *ICES Journal of Marine Science: Journal du Conseil*, 61(4):501–507, 2004.

[53] Michael R. Landry, John Constantinou, and Julie Kirshtein. Microzooplankton grazing in the central equatorial Pacific during February and August, 1992. *Deep-Sea Research II*, 42(2-3):657–671, 1995.

[54] M.R. Landry and R.P. Hassett. Estimating the grazing impacts of microzooplankton. *Marine Biology*, 67:283–288, 1982.

[55] Bikas Sinha Lawrence Lin, A.S. Hedayat and Min Yang. Statistical methods in assessing agreement: Models, issues and tools. *Journal of the American Statistical Association*, 97(457):257–270, March 2002.

[56] SJ Lentz. A climatology of salty intrusions over the continental shelf from Georges Bank to Cape Hatteras. *Journal of Geophysical Research*, 108(C10):3326, 2003.

[57] Steven J Lentz. Observations and a model of the mean circulation over the Middle Atlantic Bight continental shelf. *Journal of Physical Oceanography*, 38(6):1203–1221, 2008.

[58] William K. W. Li. Annual average abundance of heterotrophic bacteria and *Synechococcus* in surface ocean waters. *Limnology and Oceanography*, 43(7):1746–1753, 1998.

[59] William K.W. Li. Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnology and Oceanography*, 39:169–175, 1994.

[60] WKW Li and PM Dickie. Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. *Cytometry*, 44(3):236–246, 2001.

[61] Wolfgang Ludwig, Oliver Strunk, Ralf Westram, Lothar Richter, Harald Meier, Arno Buchner, Tina Lai, Susanne Steppi, Gangolf Jobb, Wolfram Förster, et al. ARB: a software environment for sequence data. *Nucleic acids research*, 32(4):1363–1371, 2004.

[62] Hongbin Lui, Lisa Campbell, and Michael R. Landry. Growth and mortality rates of *Prochlorococcus* and *Synechococcus* measured with a selective inhibitor technique. *Marine Ecology Progress Series*, 116:277–287, 1995.

[63] Nicholas H Mann. Phages of the marine cyanobacterial picophytoplankton. *FEMS microbiology reviews*, 27(1):17–34, 2003.

[64] Sophie Mazard, Martin Ostrowski, Frédéric Partensky, and David J Scanlan. Multi-locus sequence analysis, taxonomic resolution and biogeography of marine *Synechococcus*. *Environmental Microbiology*, 14(2):372–386, 2012.

[65] R.E. McDuff and S.W. Chisholm. The calculation of in situ growth rates of phytoplankton populations from fractions of cells undergoing mitosis: A clarification. *Limnology and Oceanography*, 27(4):783–788, 1982.

[66] Lisa R Moore and Sallie W Chisholm. Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnology and Oceanography*, 44(3):628–638, 1999.

[67] Lisa R. Moore, Ralf Goerickle, and Sallie W. Chisholm. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Marine Ecology Progress Series*, 116:259–275, 1995.

[68] Lisa R Moore, Anton F Post, Gabrielle Rocap, and Sallie W Chisholm. Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnology and Oceanography*, 47(4):989–996, 2002.

[69] Lisa R Moore, Gabrielle Rocap, and Sallie W Chisholm. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature*, 393(6684):464–467, 1998.

[70] J Jeffrey Morris, Robin Kirkegaard, Martin J Szul, Zackary I Johnson, and Erik R Zinser. Facilitation of robust growth of Prochlorococcus colonies and dilute liquid cultures by "helper" heterotrophic bacteria. *Applied and Environmental Microbiology*, 74(14):4530–4534, 2008.

[71] Martin Mühling, Nicholas J Fuller, Andrew Millard, Paul J Somerfield, Dominique Marie, William H Wilson, David J Scanlan, Anton F Post, Ian Joint, and Nicholas H Mann. Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. *Environmental Microbiology*, 7(4):499–508, 2005.

[72] Scott W Nixon, Stephen Granger, Betty A Buckley, Melissa Lamont, and Brenda Rowell. A one hundred and seventeen year coastal water temperature record from Woods Hole, Massachusetts. *Estuaries*, 27(3):397–404, 2004.

[73] Robert J. Olson, Sallie W. Chisholm, Eric R. Zettler, and E. V. Armbrust. Pigments, size, and distribution of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnology and Oceanography*, 35:45–58, 1990.

[74] Robert J. Olson, Alexi A. Shalapyonok, and Heidi M. Sosik. An automated submersible flow cytometer for pico- and nanophytoplankton: FlowCytobot. *Deep-Sea Research I*, 50:301–315, 2003.

[75] Robert J Olson and Heidi M Sosik. A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot. *Limnology and Oceanography: Methods*, 5:195–203, 2007.

[76] Ryan W Paerl, Rachel A Foster, Bethany D Jenkins, Joseph P Montoya, and Jonathan P Zehr. Phylogenetic diversity of cyanobacterial *narB* genes from various marine habitats. *Environmental Microbiology*, 10(12):3377–3387, 2008.

[77] Ryan W Paerl, Kenneth S Johnson, Rory M Welsh, Alexandra Z Worden, Francisco P Chavez, and Jonathan P Zehr. Differential distributions of *Synechococcus* subgroups across the California current system. *Frontiers in Microbiology*, 2:59, 2011.

[78] Carmen Palacios, Erik Zettler, Ricardo Amils, and Linda Amaral-Zettler. Contrasting microbial community assembly hypotheses: A reconciling tale from the Río Tinto. *PLoS ONE*, 3(12):e3853, 12 2008.

[79] Brian Palenik. Cyanobacterial community structure as seen from RNA polymerase gene sequence analysis. *Applied and Environmental Microbiology*, 60(9):3212–3219, September 1994.

[80] Brian Palenik. Chromatic adaptation in marine *Synechococcus* strains. *Applied and Environmental Microbiology*, 67(2):991–994, 2001.

[81] Emily E Peacock, Robert J Olson, and Heidi M Sosik. Parasitic infection of the diatom *Guinardia delicatula*, a recurrent and ecologically important phenomenon on the New England Shelf. *Marine Ecology Progress Series*, 503:1–10, 2014.

[82] Sigrid Penno, Debbie Lindell, and Anton F Post. Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environmental Microbiology*, 8(7):1200–1211, 2006.

[83] Justine Pittera, Florian Humily, Maxine Thorel, Daphné Grulois, Laurence Garczarek, and Christophe Six. Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. *The ISME Journal*, 8(6):1221–1236, 2014.

[84] Anton F Post, Sigrid Penno, Keren Zandbank, Adina Paytan, Susan M Huse, and David Mark Welch. Long term seasonal dynamics of *Synechococcus* population structure in the Gulf of Aqaba, Northern Red Sea. *Frontiers in Microbiology*, 2:131, 2011.

[85] Mario Quevedo and Ricardo Anadón. Protist control of phytoplankton growth in the subtropical north-east atlantic. *Marine Ecology Progress Series*, 221:29–38, 2001.

[86] Julie Reveillaud, Loïs Maignien, Murat A Eren, Julie A Huber, Amy Apprill, Mitchell L Sogin, and Ann Vanreusel. Host-specificity among abundant and rare taxa in the sponge microbiome. *The ISME journal*, 8(6):1198–1209, 2014.

[87] Gabrielle Rocap, Frank W Larimer, Jane Lamerdin, Stephanie Malfatti, Patrick Chain, Nathan A Ahlgren, Andrae Arellano, Maureen Coleman, Loren Hauser, Wolfgang R Hess, Zackary I Johnson, Miriam Land, Debbie Lindell, Anton F Post, Warren Regala, Manesh Shah, Stepahnie L Shaw, Claudia Steglich, Matthew B Sullivan, Claire S Ting, Andrew Tolonen, Eric A Webb, Erik R Zinser, and Sallie W Chisholm. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature*, 424(6952):1042–1047, 2003.

[88] K.M. Rodda. *Temporal and Spatial Dynamics of* Synechococcus spp. and *Micromonas pusilla* Host-Viral Systems. PhD thesis, University of Texas at Austin, TX, 1996.

[89] Shovonlal Roy and J Chattopadhyay. Towards a resolution of 'the paradox of the plankton': A brief overview of the proposed mechanisms. *Ecological complexity*, 4(1):26–33, 2007.

[90] Ben Ryall, Gustavo Eydallin, and Thomas Ferenci. Culture history and population heterogeneity as determinants of bacterial adaptation: the adaptomics of a single environmental transition. *Microbiology and Molecular Biology Reviews*, 76(3):597–618, September 2012.

[91] David J. Scanlan. *Ecology of Cyanobacteria II*, chapter Marine Picocyanobacteria. Springer, 2012.

[92] David J Scanlan, Martin Ostrowski, Sophie Mazard, Alexis Dufresne, Laurence Garczarek, Wolfgang R Hess, Anton F Post, Martin Hagemann, I Paulsen, and Frédéric Partensky. Ecological genomics of marine picocyanobacteria. *Microbiology and Molecular Biology Reviews*, 73(2):249–299, 2009.

[93] David J Scanlan and Nyree J West. Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiology Ecology*, 40(1):1–12, 2002.

[94] Patrick D Schloss, Sarah L Westcott, Thomas Ryabin, Justine R Hall, Martin Hartmann, Emily B Hollister, Ryan A Lesniewski, Brian B Oakley, Donovan H Parks, Courtney J Robinson, Jason W Sahl, Blaz Stres, Gerhard G Thanllinger, David J Van Horn, and Carolyn F Weber. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75(23):7537–7541, 2009.

[95] Daniel Sher, Jessie W Thompson, Nadav Kashtan, Laura Croal, and Sallie W Chisholm. Response of *Prochlorococcus* ecotypes to co-culture with diverse marine bacteria. *The ISME Journal*, 5(7):1125–1132, 2011.

[96] Christian K Sieracki, Michael E Sieracki, and Charles S Yentsch. An imaging-in-flow system for automated analysis of marine microplankton. *Marine Ecology Progress Series*, 168:285–296, 1998.

[97] Christophe Six, Jean-Claude Thomas, Laurence Garczarek, Martin Ostrowski, Alexis Dufresne, Nicolas Blot, David J Scanlan, and Frédéric Partensky. Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biology*, 8(12):R259, 2007.

[98] Heidi M. Sosik, Robert J. Olson, Michael G. Neubert, Alexi Shalapyonok, and Andrew R. Solow. Growth rates of coastal phytoplankton from time-series measurements with a submersible flow cytometer. *Limnology and Oceanography*, 48(5):1756–1765, 2003.

[99] Alexandros Stamatakis. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21):2688–2690, 2006.

[100] John Steele, editor. *Spatial Pattern in Plankton Communities*, volume 3 of *Nato Conference Series*. Springer, 1978.

[101] Maayke Stomp, Jef Huisman, Floris de Jongh, Annelies J Veraart, Daan Gerla, Machteld Rijkeboer, Bas W Ibelings, Ute IA Wollenzien, and Lucas J Stal. Adaptive divergence in pigment composition promotes phytoplankton biodiversity. *Nature*, 432(7013):104–107, 2004.

[102] Maayke Stomp, Jef Huisman, Lucas J Stal, and Hans CP Matthijs. Colorful niches of phototrophic microorganisms shaped by vibrations of the water molecule. *The ISME Journal*, 1(4):271–282, 2007.

[103] Maayke Stomp, Jef Huisman, Lajos Vörös, Frances R Pick, Maria Laamanen, Thomas Haverkamp, and Lucas J Stal. Colourful coexistence of red and green picocyanobacteria in lakes and seas. *Ecology Letters*, 10(4):290–298, 2007.

[104] S.L. Strom, M.A. Brainard, J.L. Holmes, and M.B. Olson. Phytoplankton blooms are strongly impacted by microzooplankton grazing in coastal North Pacific waters. *Marine Biology*, 138:355–368, 2001.

[105] Vera Tai, Ronald S Burton, and Brian Palenik. Temporal and spatial distributions of marine *Synechococcus* in the Southern California Bight assessed by hybridization to bead-arrays. *Marine Ecology Progress Series*, 426:133–147, 2011.

[106] Vera Tai and Brian Palenik. Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *The ISME journal*, 3(8):903–915, 2009.

[107] Koichiro Tamura, Joel Dudley, Masatoshi Nei, and Sudhir Kumar. Mega4: molecular evolutionary genetics analysis (mega) software version 4.0. *Molecular biology and evolution*, 24(8):1596–1599, 2007.

[108] Gerardo Toledo, B Palenik, and B Brahamsha. Swimming marine *Synechococcus* strains with widely different photosynthetic pigment ratios form a monophyletic group. *Applied and Environmental Microbiology*, 65(12):5247–5251, 1999.

[109] An-Yi Tsai, Kuo-Ping Chiang, Jeng Chang, and Gwo-Ching Gong. Seasonal variations in trophic dynamics of nanoflagellates and picoplankton in coastal waters of the western subtropical pacific ocean. *Aquatic microbial ecology*, 51(3):263–274, 2008.

[110] An-Yi Tsai, Gwo-Ching Gong, Robert W Sanders, Kuo-Ping Chiang, Jun-Kai Huang, and Ya-Fan Chan. Viral lysis and nanoflagellate grazing as factors controlling diel variations of *Synechococcus* spp. summer abundance in coastal waters of Taiwan. *Aquatic Microbial Ecology*, 66(2):159–167, 2012.

[111] Daniel Vaulot and Domninique Marie. Diel variability of photosynthetic picoplankton in the equatorial Pacific. *Journal of Geophysical Research*, 104(C2):3297–3310, February 1999.

[112] John B Waterbury and Frederica W Valois. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Applied and Environmental Microbiology*, 59(10):3393–3399, 1993.

[113] John B Waterbury, Stanley W Watson, Robert RL Guillard, and Larry E Brand. Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature*, 277:293–294, 1979.

[114] John B Waterbury, Stanley W Watson, Frederica W Valois, and Diana G Franks. Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. *Canadian Bulletin of Fisheries and Aquatic Sciences*, 214:71–120, 1986.

[115] John B Waterbury, Joanne M Willey, Diana G Franks, Frederica W Valois, and Stanley W Watson. A cyanobacterium capable of swimming motility. *Science*, 230(4721):74–76, 1985.

[116] Nyree J West and David J Scanlan. Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Applied and Environmental Microbiology*, 65(6):2585–2591, 1999.

[117] A. Michelle Wood. Discrimination between types of pigments in marine *Synechococcus* spp. by scanning spectroscopy, epifluorescnece microscopy, and flow cytometry. *Limnology and Oceanography*, 30(6):1303–1315, 1985.

[118] Alexandra Z. Worden and Brian J. Binder. Application of dilution experiments for measuring growth and mortality rates among *Prochlorococcus and* Synechococcus populations in oligotrophic environments. *Aquatic Microbial Ecology*, 30:159–174, 2003.

[119] Erik R. Zinser and Roberto Kolter. Mutations enhancing amino acid catabolism confer a growth advantage in stationary phase. *Journal of Bacteriology*, 181(18):5800–5807, 1999.

[120] Katrin Zwirglmaier, Ludwig Jardillier, Martin Ostrowski, Sophie Mazard, Laurence Garczarek, Daniel Vaulot, Fabrice Not, Ramon Massana, Osvaldo Ulloa, and Dave J Scanlan. Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environmental Microbiology*, 10(1):147–161, 2008.

[121] Katrin Zwirglmaier, Ed Spence, Mikhail V Zubkov, David J Scanlan, and Nicholas H Mann. Differential grazing of two heterotrophic nanoflagellates on marine *Synechococcus* strains. *Environmental Microbiology*, 11(7):1767–1776, 2009.