The Growth and Activity of Genetically Diverse Prochlorococcus

by

Yajuan Lin

Marine Science and Conservation
Duke University

Date: __________________________
Approved: ________________________

___________________________
Zackary Johnson, Supervisor

___________________________
Erik Zinser

___________________________
Dana Hunt

___________________________
Jennifer Wernegreen

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy
in Marine Science and Conservation
in the Graduate School
of Duke University

2013
ABSTRACT

The Growth and Activity of Genetically Diverse Prochlorococcus

by

Yajuan Lin

Marine Science and Conservation
Duke University

Date: _____________________

Approved:

___________________________
Zackary Johnson, Supervisor

___________________________
Erik Zinser

___________________________
Dana Hunt

___________________________
Jennifer Wernegreen

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Marine Science and Conservation in the Graduate School of Duke University

2013
Abstract

While much is known about the abundance and genetic diversity of environmental microbial communities, little is known about taxon-specific activity. In this thesis I address this gap using a model marine microbe, the cyanobacterium *Prochlorococcus*, which numerically dominates surface water of tropical and subtropical open oceans and encompasses a group of genetically defined clades that are ecologically distinct. Ribosomal RNA is a promising indicator of *in situ* activity because of its essential role in protein synthesis as well as its phylogenetic information, which could be used to distinguish clades among mixed populations. Here I show that, in a cultured model system that the specific growth rate of representative *Prochlorococcus* strains could be quantitative predicted from cellular rRNA content (assessed by RT-qPCR), cell size (assessed by flow cytometry) and temperature. Applying this approach in the field, I observed unique clade-specific activity patterns for *Prochlorococcus*. For example, vertically within the euphotic zone, eHL-II activity is strongly impacted by light and is consistent with photosynthesis and on a horizontal transect from Hawaii to San Diego, eHL-I and eHL-II activities exhibit significant transitions and appear to be regulated by temperature, nutrient and vertical mixing gradients. Using ribosomal tag pyrosequencing of DNA and RNA, I have extended our observation to the Eubacterial
community and described the biomass distribution (rDNA) and activity (rRNA) patterns from two representative depths (25 and 100 m) at a well-studied oligotrophic ocean station. These results show that for some populations the abundances and activities are significantly uncoupled, which suggests substantial top-down controls or physical transport processes. Further exploring the taxon-specific activity patterns along with abundances and environmental variables across time and space is essential to better understanding the dynamics of a complex microbial system as well as predicting the consequences of environmental change.
# Contents

Abstract ......................................................................................................................................... iv

List of Tables................................................................................................................................. ix

List of Figures ................................................................................................................................ x

1. Introduction ............................................................................................................................... 1
   1.1 Microbial activity: methods and ecology ................................................................. 3
   1.2 *Prochlorococcus* a model organism of global significance................................. 8
   1.3 Explore the clade specific activity of *Prochlorococcus* in the ocean ...................... 14

2. *In situ* activity of a dominant *Prochlorococcus* ecotype (eHL-II) from rRNA content and cell size ................................................................................................. 19
   2.1 Introduction................................................................................................................... 19
   2.2 Experimental Procedures .......................................................................................... 24
     2.2.1 Culture conditions and sampling ........................................................................ 25
     2.2.2 Field sample collection and characterization ...................................................... 26
     2.2.3 DNA extraction and quantitative PCR................................................................. 27
     2.2.4 RNA extraction and RT-qPCR .............................................................................. 27
     2.2.5 Flow cytometry .................................................................................................... 28
     2.2.6 Photosynthesis (Primary production) .................................................................. 29
   2.3 Results and discussion ................................................................................................. 30
     2.3.1 Relationship of rRNA and cell size to specific growth rate ................................. 30
     2.3.2 Diel variation ........................................................................................................ 35
     2.3.3 Field patterns and relationship to environmental variables ............................... 38
4.2.3 Primary production (photosynthesis activity) ......................................................76
4.2.4 Electron transport system (ETS) activity..........................................................77
4.2.5 ATP .....................................................................................................................78
4.2.6 RNA extraction and RT-qPCR .............................................................................79
4.2.7 Field sample collection and environmental characterization .............................80
4.3 Results and discussion ..........................................................................................81
4.3.1 Culture growth ....................................................................................................81
4.3.2 Activity Metrics as a Function of Temperature ..................................................83
4.3.3 Relationship of rRNA content and growth rate ..................................................86
4.3.4 Field patterns across a natural sea surface temperature gradient .....................90
4.4 Conclusions .............................................................................................................96
Conclusions ..................................................................................................................98
Appendix .......................................................................................................................106
References ....................................................................................................................108
Biography ......................................................................................................................130
List of Tables

Table 1: Summary of Prochlorococcus growth rate studies from the field............................ 13

Table 2: Statistical parameters of the linear regressions shown in Fig. 3B-F ...................... 32
List of Figures

Figure 1: (A) Phylogenetic relationship of Prochlorococcus isolates inferred by Neighbor-Joining tree from Internal Transcribed Spacer (ITS). These isolates fall into two distinct clades that are high- and low-light adapted. These groups can be further subdivided into six ecologically distinct clades including eHL-II (red), eHL-I (green), eLL-III (orange), eLL-II (purple), eLL-I (dark blue), and eLL-IV (light blue) (Ahlgren et al 2006). (B) Typical vertical profiles of the abundance of each ecotype in the North Pacific Subtropical Gyre (eLL-I and eLL-II are below the detection limit) .................................................. 11

Figure 2: (A) Specific growth rate ($\mu$) as a function of irradiance and (B) side scatter (SSC) as index of cell size, (C) rRNA cell$^{-1}$, and (D) rRNA cell$^{-1}$ normalized by SSC for MIT9312 (open circles) and MIT9215 (closed circles) as a function of growth rate. Mean rRNA cell$^{-1}$ SSC$^{-1}$ vs. growth rate of the diel experiment is shown as gray square in (D). 33

Figure 3: Diel variation in Prochlorococcus strain MIT9312 of (A) cell concentration, (B) side scatter (SSC), (C) total population 23S rRNA, (D) rRNA cell$^{-1}$ and (E) rRNA cell$^{-1}$ SSC$^{-1}$ (closed triangles) and relative irradiance (solid line). Shaded area represents incubator dark period ........................................................................................................................................................................ 36

Figure 4: Vertical profiles of $^{14}$C based primary production (open symbols) and total eHL-II rRNA (closed symbols) for stations in (A) the North Pacific Subtropical Gyre (NPSG) (Station #1, circles) and (B) the Western Equatorial Pacific (Station #2, squares) and (C) The relationship between rRNA and photosynthesis (carbon uptake) (circles – Station #1, squares – Station #2). .................................................................................................................................................. 39

Figure 5: Vertical profiles of (A, E) light (% surface irradiance), temperature ($^\circ$C) and nitrate + nitrite (N+N, $\mu$M), (B, F) eHL-II and heterotrophic bacterioplankton abundance (cells ml$^{-1}$), (C, G) side scatter (SSC) of Prochlorococcus relative to 1.0 $\mu$m polystyrene beads and (D, H) specific activity of eHL-II (rRNA cell$^{-1}$ SSC$^{-1}$) for representative locations in a moderate mixing site (A-D) and a deep mixing site (E-H) in the North Pacific Subtropical Gyre .......................................................................................................................................................... 44

Figure 6: Vertical variability of key physical, chemical and biological variables at station ALOHA. (A) Irradiance (% of surface) and nitrate ($\mu$mol kg$^{-1}$). Arrows indicate the depths of samples chosen for bacterioplankton diversity analysis (B) Chlorophyll a (mg m$^{-3}$) and temperature (C). (C) Cellular abundance measured by flow cytometry for Prochlorococcus and non-photosynthetic bacterioplankton (cells mL$^{-1}$) ............................................................................................................. 58
Figure 7: Rarefaction curves for 16S rDNA and rRNA libraries from station ALOHA at 25 m and 100 m depth. OTUs were defined at 97% identity to minimize sequencing errors.

Figure 8: Shared and unique OTUs (97% identity) for 16S rDNA and rRNA libraries from station ALOHA at 25 m and 100 m depth.

Figure 9: Phylogenetic relationship of bacterial diversity with activity and abundance measurements for 16S rRNA and rDNA libraries from station ALOHA at 25 m and 100 m. (A) Maximum likelihood phylogenetic tree of the partial 16S rRNA sequences for the 50 most abundant OTUs (97% identity) in equally sub-sampled 16S rRNA and rDNA libraries (75,000 members). Bootstrap percentages greater than 80 for a given branch are indicated by a small circle. (B) Heat map showing the log abundance (log (observations+1)) for each OTU from the tree in (A) for RNA and DNA libraries. (C) Ratio of RNA to DNA (rRNA/ (rDNA+1)) showing the specific activity of each OTU for libraries from 25 and 100 m.

Figure 10: Relationship between activity (log rRNA+1) and abundance (log rDNA+1) at 25 m (open squares) and 100 m (closed circles) of each OTU. To eliminate bias, correlations are for log rRNA and log rDNA and therefore limited to OTUs where both rRNA and rDNA was present.

Figure 11: Prochlorococcus abundance and activity with associated physicochemical variability versus depth for two representative locations in the Equatorial Pacific (A-C) and Sargasso Sea (D-F). Abundance profiles (A,D) of the two dominant ecotypes high light eMIT9312/eHL-II and low light eMIT9313/eLL-IV measured using qPCR of 23S rRNA genes (cells mL$^{-1}$). Normalized specific activity (B,E) of eHL-II and eLL-IV ecotypes measured using qPCR of 23S rRNA and rDNA. Profiles of physicochemical variability (C,F) including irradiance (% of surface), temperature (°C) and nitrate (µmol kg$^{-1}$).

Figure 12: (A) Specific growth rate (d$^{-1}$) as a function of temperature, and (B) side scatter as a function of Temperature for MIT9312 (red triangles) and MED4 (blue circles).

Figure 13: Variation of cellular activities (A) primary production per cell, (B) rRNA per cell, (C) ATP (pmol) per cell and (D) ETS activity per cell across a gradient of temperature for MIT9312 (red triangles) and MED4 (blue circles).
Figure 14: (A) rRNA cell⁻¹, and (B) rRNA cell⁻¹ normalized by SSC as a function of growth rate for MIT9312 (red triangles), MED4 (blue circles) and previous light experiment results (black crosses).

Figure 15: (A) Sample locations and sea surface temperature (SST in °C) in the North Pacific Ocean. Surface transect profiles of (B) biomass distribution, (C) rRNA per cell for Prochlorococcus ectypes eHL-II or eMIT9312 (red triangles) and eHL-I or eMED4 (blue circles), and (D) environmental variables including SST, PO4 (red dashed line), mixed layer depth (black squares) and percentage of small size phytoplankton chlorophyll (0.2 – 2 µm). Grey arrows highlight two key transition points.

Figure 16: Potential growth rates calculated from Equation (2) for eHL-II or eMIT9312 (red) and eHL-I or eMED4 (blue) along a latitudinal sea surface temperature gradient (black dotted line). Mixed layer depth is shown as solid black line.

Figure 17: Vertical profiles of (A) light (% surface PAR) and temperature (°C), (B) eHL-II (eMIT9312), eLLIV (eMIT9313) and total Prochlorococcus abundance (cells ml⁻¹), (C) rRNA cell⁻¹ for eHLII and eLLIV, and (D) rRNA cell⁻¹ normalized by SSC. Arrows point to potentially viable deep Prochlorococcus.

Figure 18: Preservation and lysis methods comparison. A) DNA extraction and 23S gene qPCR. B) RNA extraction and 23S gene RT-qPCR. Three RNA preservation solutions are RLT (Qiagen) with 1% 2-mercaptoethanol, RNA Shield (Zymo Research) and RNAlater (Ambion). Lighter colored bars are treated with bead-beating. Arrow points to the default RNA or DNA extraction method. Asterisks indicate values calculated from 100-fold diluted samples which cannot be amplified at original concentration due to qPCR inhibition.
Acknowledgements

I would like to thank my advisor, Zackary Johnson for his mentorship in scientific thinking, sharing his experience at sea, great patience when I came up with problems, encouragement when I struggled, and always keeping an open mind which is especially important to me as an international student. Thanks to Dana Hunt for helping me with bioinformatics and microbial ecology; and also thank her for building a diverse microbial research community in Duke Marine lab. Thanks to Jennifer Wernegreen and Erik Zinser for insightful advices and constructive comments on my projects.

Mahalo to my instructors in University of Hawaii, Mark Merrifield, Telu Li, Craig Smith, and Matt Church. They led me to this fascinating oceanography world. And thanks to my colleague PhD students and Postdocs for great companionship and inspiring discussion. In particular, I’m grateful to Thais Bittar, Maria de Oca, Alyse Larkin, Chris Ward, Charmaine Yung, Anna Ritchie, Lynne Williams and Sophie Plouviez. They are great colleagues and wonderful friends.

Marine lab is a small and sweet community. Thanks to Rachel, Patty, Helen, Tom, Jeff, and Janil. They were always there when I needed help.

Finally, special thanks to my family. I would have not come this far without their love and support.
This work was made possible with financial support from:

The Duke Marine Lab and Duke University

University of Hawaii at Manoa

NSF to Zackary Johnson (#0959630, 1031064 and OCE10-31064)

DOE Joint Genome Institute
1. Introduction

“Microbes are the foundation for all of life. From the air we breathe to the soil we rely on for farming to the water we drink, everything humans need to survive is intimately coupled with the activities of microbes”.

-The Uncharted Microbial World (Harwood and Buckley 2007)

Microorganisms are the most abundant form of life on the planet. Their diverse metabolic activities enable them to thrive in a wide spectrum of environments. Characterized by high abundance and small size, which affords them a proportionally large surface area, they act as efficient biochemical machines to transform compounds and energy. For example, several key microbial members serve in natural environments as, 1) decomposers – recycling organic matters and nutrients, 2) photosynthesizers – capturing solar energy and fixing carbon, 3) nitrogen fixers – generate N flow into the biosphere, 4) chemolithoautotrophs in the deep sea – utilize redox energy from minerals to sustain the hydrothermal vent ecosystem. Characterizing and quantifying microbial activities is essential to developing an understanding of how the microbial community impacts the environment (functionality) and responds to external changes (dynamics), to parameterizing the fluxes associated with biogeochemical cycles, and in applied microbiology applications such as identifying key (i.e. highly active) species of ecological, industrial and clinic interests.
One major challenge associated with environmental activity studies is that the natural microbial communities are highly diverse (both genetically and physiologically) and heterogeneous systems. If phylogenetically diversified clades (or ecotypes if evidence of ecological differentiation) react to environmental variables differentially, then as a whole the total community response to the environment will be difficult to predict because of differential responses of individual clades. Even between genetically identical cells there could be heterogeneity in their activities due to different micro-scale environmental gradients that they experience. How can the contribution of genetically distinct cells be assessed? And what phylogenetic resolution is sufficient to capture patterns of activity that are ecologically and biogeochemically relevant?

Regardless of potential technical challenges, culture-independent methods to study activities in situ are of particular interest because the majority of microbial life is still uncultivated. And the complex interactions between microbes can only be revealed in context of other community members. Further, the natural environmental variations across time and space are difficult to reconstruct in lab without oversimplification or human artifacts.
1.1 Microbial activity: methods and ecology

Recognizing its importance, several approaches have been developed to quantify microbial activity by quantifying different metabolic processes or cell growth rates. (Del Giorgio and Gasol 2008) categorized different activity measurements at the cellular level based on cell physiology and function and categorized them into nine major groups: (1) morphological integrity; (2) DNA quantification and status; (3) respiration activity; (4) the status of the membrane; (5) the internal enzymatic capacity; (6) the uptake of organic substances or CO₂, O₂ exchange; (7) the synthesis of DNA; (8) cell growth; (9) the amount of rRNA. Those activities focus on different aspects of cell metabolism or growth, but within a cell as a whole functional unit they are intercorrelated with each other through energy and element flux. For example, for a heterotroph cell the net growth is a result of total organic substance incorporation minus loss through respiration or fermentation.

Among those different metrics of activity, biomass production is of intrinsic interest to both microbial ecologists and biogeochemists because it is related to cell growth and elements (such as carbon) flux into the system. The most commonly applied methods for estimating environmental community production is radioisotope labeling – after incubated with a radioactive tracer, substance assimilation is assessed through radioactivity then converted to biomass production. For example, ³H-labeled thymidine (TdR), incorporated into cellular DNA, was first introduced about 30 years ago.
(Fuhrman and Azam 1982; Fuhrman and Azam 1980), and later an alternative approach was developed based on incorporation of \(^3\)H-labeled leucine into bacterial protein (Kirchman et al. 1985a; Simon and Azam 1989). The former is related to cell replication, while the latter targets general protein accumulation. Both methods have been intensively applied in aquatic environments to estimate heterotroph bacterial community productivity (Ducklow 2000; Ducklow et al. 1995; Kirchman et al. 1993; Pérez et al. 2009; Shish and Ducklow 1994). Similarly phytoplankton production is usually assessed from C\(^{14}\) labeled CO\(_2\) incorporation (Steemann 1952), change of O\(_2\) concentration (Gaarder and Gran 1927), or the oxygen produced from O\(^{18}\) labeled H\(_2\)O (Grande et al. 1989). One additional advantage in photosynthetic organism by combining light and dark incubations, different activity components (i.e. gross production, net production and respiration) can be teased apart (Ducklow 2000).

Moreover, with accumulated knowledge of photosynthesis mechanism combined with modern remote sensing technology, global scale production can be estimated from satellite derived chlorophyll (standing stock), light and temperature (Behrenfeld et al. 2006; Field et al. 1998). This approach is extremely powerful in data handling and observing range (temporal and spatial); however, it does not provide any detail on the variability in activity among the different populations within the community.

More recently activity measurements have been made utilizing high-throughput molecular approaches to assess the expression pattern of RNA or proteins as indirect
indicators for cellular activities. These techniques include 1) microarrays (Tolonen et al. 2006; Zinser et al. 2009), 2) metatranscriptomics including mRNA enriched cDNA library (Frias-Lopez et al. 2008; Shi et al. 2010) and library specifically targeting ribosomal RNA (Campbell et al. 2011b; Placella et al. 2012), and 3) metaproteomics (Verberkmoes et al. 2008; Williams et al. 2012). These methods have the potential to examine natural microbial community activity in great depth. For example, a 16S rRNA amplicon pyrosequencing survey of a lake (Jones and Lennon 2010b) reveals that many rare bacterial taxa, which cannot be characterized by traditional sanger sequencing, are disproportionally active in the community. Moreover, these methods can assess different metabolic activities simultaneously thus enable researchers to compare and identify the key microbe mediated biochemical activities within an ecosystem. A metaproteomic study of the bacterioplankton community from the Antarctic coastal waters (Williams et al. 2012) reveals that in summer oxygenic photoautotrophy performed by phytoplankton drives the major carbon fixation, while in winter chemoautotrophic carbon fixation dominates the system and is led by ammonia-oxidizing archaea and nitrite-oxidizing bacteria. These techniques are innovative and powerful because of their taxonomic resolution and ability to characterize diverse activities in natural environments, which was previously not possible. However, the fundamental assumption of these approaches is that there is a quantitative correlation between the expression of RNA or protein and specific activity. This assumption has
been tested in some model organisms such as relating rRNA to growth rate in *Escherichia coli* and *Salmonella typhimurium* (Bremer and Dennis 1996b; Delong et al. 1989; Dortch et al. 1983; Schaechter et al. 1958). However, there are only limited numbers of cases reported from environmental isolates, and no general relationship has been concluded (Flynn et al. 2010). The investigations of rRNA as specific activity indicator will be further reviewed and discussed in later part of this introduction.

Since activity is relatively difficult to assess in the field, one intuitive solution is using the abundance of a specific taxa to indicate its activity. However, abundance and activity are related but not necessarily coupled. Population abundance represents the time-integrated result of cell number changing rates, which is controlled by the balance between cell growth (a type of activity, and related to various physiological activities) and removal rate. In a simplified system, assuming steady state and even removal rate, the faster growing population accumulates more biomass and becomes more abundant with time. The positive correlation between population activity (defined as growth here) and abundance has been reported as a common trend in microbial communities from coral sediments and surface coastal waters (Campbell and Kirchman 2012; Campbell et al. 2011b; Gaidos et al. 2011). Under more realistic conditions the removal rate, which is not constant and controlled by a combination of several processes including grazing, virus lysis and cell death, can lead to negative correlations between abundance and activity through indirect mechanisms. For example, protists have been
reported selectively graze on more active bacteria (Del Giorgio et al. 1996; Landry et al. 1991; Massana et al. 2009; Tadonleke et al. 2005), and in the “kill the winner” hypothesis (Thingstad 2000; Thingstad and Lignell 1997) viruses can effectively remove the most actively growing bacterial host. When these processes overpower the direct positive correlation between activity and abundance, activity and abundance can appear uncoupled or even negatively related. Other processes can lead to departures in the system from steady state, such as physical mixing or nutrient pulse, which also can disrupt the relationship between activity and abundance. Indeed in the few studies to date, it appears that activity and abundance are uncoupled (Campbell et al. 2011b; Jones and Lennon 2010b).

Activity provides unique information about the dynamics and ecological functions of microbial community, and it is related to abundance through direct or indirect processes. Exploring microbial activity in combination with abundance and diversity can help us to further understand the mechanisms driving microbial ecosystems.
1.2 Prochlorococcus a model organism of global significance

The single-celled cyanobacterium *Prochlorococcus* numerically dominates open-ocean ecosystems, and plays a critical role in marine microbial ecology and biogeochemical cycles. Characterized by its small size (0.5 - 0.8 µm diameter) and unique pigment content (divinyl chlorophyll a and b), *Prochlorococcus* has been found to be ubiquitously distributed in tropical and subtropical oceans (Campbell et al. 1998; Campbell et al. 1997; Chisholm et al. 1988; Johnson et al. 2006; Rusch et al. 2010; West et al. 2010; Yooseph et al. 2010; Zubkov et al. 2000b; Zwirglmaier et al. 2007). In a typical vertical profile where it is abundant (e.g. Fig. 1B), *Prochlorococcus* cell concentrations are ~ 10^5 cells ml\(^{-1}\) from the surface to the base of the mixed layer, then decrease to below 100 cells ml\(^{-1}\) at the base of the euphotic zone (~200 m). Its ecological success has been attributed in part to its small cell size which helps to minimize its resource requirement (Bertilsson et al. 2003), avoid large grazers (Landry et al. 1997), and maximize the nutrient and light acquisition (Chisholm et al. 1992; Raven 1998). In open-ocean ecosystems, this organism is estimated to contribute to 13% - 48% of the total primary production (Goericke and Welschmeyer 1993; Vaulot et al. 1995), and it is responsible for significant amount of energy and carbon flux into food web and biogeochemical cycles (Barber 2007; Landry 2002; Richardson and Jackson 2007).
One hypothesis to explain the dominance of Prochlorococcus across large environmental gradients (e.g. from the surface to the base of euphotic zone) is that within this genus phylogenetically and physiologically diverse clades can co-exist and occupy different ecological niches (Moore et al. 1998; Rocap et al. 2002). Phylogenetic analyses using \textit{rpoC1}, 16S rRNA and ITS (Internal Transcribed Spacer) sequences have revealed distinct evolutionary lineages within this group (Palenik and Haselkorn 1992; Rocap et al. 2002) (Fig. 1A). Moreover, genomic studies further confirmed the phylogenetic diversification and suggested that different lineages have experienced frequent gain and loss of genes associated with light acclimation, nutrient assimilation, DNA repair, and membrane synthesis etc., which may make allow them to adapted to different habitats (Kettler et al. 2007; Rocap et al. 2003). Culture studies with isolates representing different genetic clades demonstrated distinct physiological characteristics such as pigment content (Moore and Chisholm 1999; Moore et al. 1998), photophysiology (Moore and Chisholm 1999; Moore et al. 1995; Moore et al. 1998; Zinser et al. 2007), temperature responses (Johnson et al. 2006; Zinser et al. 2007), and nutrient utilization (Martiny et al. 2006; Martiny et al. 2009a; Moore et al. 2002; Tolonen et al. 2006). Moreover, clade-specific qPCR primers (Ahlgren et al. 2006b; Ahlgren et al. 2006c) and oligonucleotide probes (West and Scanlan 1999) have been developed to assess the abundance of different clades in the field. Application of qPCR and dot blot hybridization techniques reveals that these Prochlorococcus clades have distinct
ecological distributions both versus depth (Fig. 1B) and across ocean basins (Johnson et al. 2006; West et al. 2010; Zwirglmaier et al. 2007), and therefore these genetic clades are considered as ecotypes (eHL-I, eHL-II, eHNLC1&2, eLL-I, eLL-II, eLL-III, and eLL-IV). The combination of laboratory based physiology and genetic studies along with biogeographic field studies show that these genetic clades have unique ecologies with important implications for ecosystem function (Follows et al. 2007; Rabouille et al. 2007).

While much is known about the genetic diversity and biomass distribution of *Prochlorococcus*, to assess how these ecotypes function differently in the ecosystem, it is necessary to measure their *in situ* activity. This term “*in situ* activity” can be alternately defined as primary production, specific growth rate, cellular ATP synthesis or others. But regardless of how it is defined, activity is the key to understanding their ecological function. While standing stock helps to the answer ‘where are they’, activity helps to understand ‘what are they doing’ in the environment. Particularly for *Prochlorococcus*, as a globally important primary producer and model prokaryote, the *in situ* growth rate and photosynthesis activity are of intrinsic interest to researchers.
Figure 1: (A) Phylogenetic relationship of Prochlorococcus isolates inferred by Neighbor-Joining tree from Internal Transcribed Spacer (ITS). These isolates fall into two distinct clades that are high- and low-light adapted. These groups can be further subdivided into six ecologically distinct clades including eHL-II (red), eHL-I (green), eLL-III (orange), eLL-II (purple), eLL-I (dark blue), and eLL-IV (light blue) (Ahlgren et al. 2006). (B) Typical vertical profiles of the abundance of each ecotype in the North Pacific Subtropical Gyre (eLL-I and eLL-II) are below the detection limit.

Traditional activity measurements in the field distinguish phytoplankton by size or pigment, and these methods treat Prochlorococcus as a single population. Several metrics of activity have been developed and include (1) size-fractionated primary production measurement through carbon isotope incorporation (Marañón et al. 2000; Moran et al. 2004), (2) selective metabolic inhibitors (Liu et al. 1995), (3) primary production measurement through pigment labeling (Goericke and Welschmeyer 1993; Redalje 1983), (4) cell division frequency through diel flow cytometer analysis (Vaulot et
al. 1995), and (5) Landry-Hassett dilution experiment (Landry and Hassett 1982; Landry et al. 1995b). Although size-fractionated primary production has been widely applied in the field, it is difficult to differentiate between Prochlorococcus and Synechococcus with certainty based only on size thus Prochlorococcus primary production is usually not separately estimated but included in the picoplankton fraction (0.2 – 2 µm) (Marañón et al. 2001; Moreno-Ostos et al. 2011; Pérez et al. 2006; Poulton et al. 2006). Diel flow cytometry and dilution experiments are the two most commonly applied methods to assess Prochlorococcus activity in the field, and some of these results are summarized in Table 1. Based on these studies and others, some key insights on Prochlorococcus activities have been revealed: (1) Prochlorococcus can grow exceptionally fast in oligotrophic oceans, with maximum growth rate reported up to 1.0 d\(^{-1}\) at upper water column and low but detectable growth at the bottom of euphotic zone; (2) Specific growth rate or production is strongly regulated by light (PAR) and vertical mixing, while macronutrients appear to have less of an impact; (3) A subsurface peak of growth rate is commonly observed within the mixed layer, and the peak depth varies from 10-100m; (4) Activity patterns are dynamic both seasonally and geographically. However, while much has been revealed about Prochlorococcus ecology using these approaches, there are two major drawbacks for these methods as they require frequent sampling-intensive labor or are potentially affected by incubation artifacts (Venrick et al. 1977). Nevertheless, these studies are limited to examining the patterns of the whole
community even though its component populations are comprised of different ecotypes.

Thus the in situ activity of specific Prochlorococcus ecotypes remains unknown.

Table 1: Summary of Prochlorococcus growth rate studies from the field

<table>
<thead>
<tr>
<th>Method</th>
<th>Location</th>
<th>Maximum growth rate (day(^{-1}))</th>
<th>Depth (m)</th>
<th>Season</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diel FCM</strong></td>
<td>Equatorial Pacific</td>
<td>0.7-0.8, 0.7-1.0</td>
<td>30, 30</td>
<td>Spring, Fall</td>
<td>(Vaulot et al. 1995)</td>
</tr>
<tr>
<td></td>
<td>Equatorial Pacific</td>
<td>0.5-0.6</td>
<td>15, 30</td>
<td>Spring, Fall</td>
<td>(Binder et al. 1996)</td>
</tr>
<tr>
<td></td>
<td>Tropical Pacific 12(^\circ) N, 12(^\circ) S,</td>
<td>0.4-0.5</td>
<td>40-100</td>
<td>El Niño &amp; normal upwelling</td>
<td>(Liu et al. 1999)</td>
</tr>
<tr>
<td></td>
<td>Equatorial Pacific 5(^\circ) N - 5(^\circ) S</td>
<td>0.5-0.6</td>
<td>10-30</td>
<td>Normal upwelling</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Equatorial Pacific 5(^\circ) N - 5(^\circ) S</td>
<td>0.5</td>
<td>10-30</td>
<td>El Niño</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Northern Atlantic gyre</td>
<td>0.14 ± 0.03</td>
<td>Surface*</td>
<td>Apr - May</td>
<td>(Zubkov et al. 2000a)</td>
</tr>
<tr>
<td></td>
<td>Equatorial Atlantic</td>
<td>0.14 ± 0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Southern Atlantic gyre</td>
<td>0.17 ± 0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dilution</strong></td>
<td>Equatorial Pacific</td>
<td>0.22-0.26</td>
<td>10, 40*</td>
<td>Aug - Sep</td>
<td>(Landry et al. 1995a)</td>
</tr>
<tr>
<td></td>
<td>Sargasso Sea</td>
<td>0.44-0.76</td>
<td>50*</td>
<td>Winter</td>
<td>(Worden and Binder 2003a)</td>
</tr>
<tr>
<td></td>
<td>Sargasso Sea</td>
<td>0.40-0.52</td>
<td>50*</td>
<td>Summer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gulf Stream</td>
<td>0.58</td>
<td>50*</td>
<td>Summer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>California Current</td>
<td>0.32-0.45</td>
<td>12, 25*</td>
<td>Summer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hawaiian Island, cold-core eddy</td>
<td>0.8-0.9</td>
<td>50</td>
<td>In Cyclone</td>
<td>(Landry et al. 2008)</td>
</tr>
<tr>
<td></td>
<td>Equatorial Pacific</td>
<td>0.6-1.0</td>
<td>10-50</td>
<td>Sep, Dec</td>
<td>(Selph et al. 2011b)</td>
</tr>
<tr>
<td><strong>Pigment labeling</strong></td>
<td>Sargasso Sea</td>
<td>0.3</td>
<td>20% surface Irradiance</td>
<td>Summer average</td>
<td>(Goericke and Welschmeyer 1993)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3</td>
<td></td>
<td>20% surface Irradiance</td>
<td>Winter average</td>
</tr>
<tr>
<td><strong>Metabolic inhibitor</strong></td>
<td>Station ALOHA</td>
<td>0.59±0.19</td>
<td>25</td>
<td>Oct</td>
<td>(Liu et al. 1995)</td>
</tr>
</tbody>
</table>

*Single depth measurement, not depth profile
1.3 Explore the clade specific activity of Prochlorococcus in the ocean

Studies from the cultured isolates and field surveys have revealed that Prochlorococcus is genetically and physiologically diverse. Different ecotypes of Prochlorococcus respond to environmental factors differently in the lab and have been observed in distinct biogeographic distributions in the field. Comparative genomic studies have revealed that different factors such as light, nutrients and temperature shaped the ecotype diversification at different degrees. Now ecotype specific activity in the field is the crucial information yet still lacking to connect existing knowledge and further understand their ecology and contribution to biogeochemical cycles. Thus, the major goal of this work is to establish and test a system to quantify Prochlorococcus ecotype specific activities in the ocean and to investigate some of the environmental drivers that may influence activity in situ.

To achieve this goal we used several approaches. First, we developed rRNA as an in situ activity probe for a dominant Prochlorococcus ecotype eHL-II in the lab and verified this technique in field stations with a diverse microbial community (Chapter 2). Ribosomal RNA is particularly well suited as an activity indicator due to its key function in general protein synthesis and the unique phylogenetic information it contains. The positive correlation between rRNA and cell specific growth rate have been intensively studied and parameterized in the model bacteria Escherichia coli and
Salmonella typhimurium (Bremer and Dennis 1996b; Delong et al. 1989; Dortch et al. 1983; Poulsen et al. 1993; Schaechter et al. 1958). More recently, taxon-specific correlations between rRNA content and cell specific growth rate have been reported in cultured environmental microbes including slow growing marine heterotrophs (Kemp et al. 1993), and marine cyanobacteria (Binder and Liu 1998; Lepp and Schmidt 1998; Parrott and Slater 1980; Worden and Binder 2003b). Using Prochlorococcus as a model organism, in this work we specifically investigated the relationship between rRNA content and cell specific growth rate for two representative strains from ecotype eHL-II. Ecotype eHL-II is a high-light adapted clade and the dominant Prochlorococcus type in the mixed layer of low and mid-latitude (<30 °C) oligotrophic oceans (Johnson et al. 2006). qPCR primers designed in 23S rRNA gene of this ecotype have been shown to successfully distinguish it from other environmental microbes (Ahlgren et al. 2006b). In our work, we cultured two representative strains of this ecotype and manipulated the irradiance level, which is the primary factor shaping the evolution and distribution of this ecotype (Martiny et al. 2009b; Rocap et al. 2003), to obtain a series of growth rate and quantified their correlation with cellular rRNA content. Then we applied this technique to representative field stations in the Pacific Ocean and successfully generated eHL-II specific activity depth profiles which are consistent with the total Prochlorococcus production patterns assessed using traditional carbon isotope incorporation method.
After observing unique rRNA activity patterns along depth using *Prochlorococcus* eHL-II as a model system, further questions are - how does the activity change between different *Prochlorococcus* ecotypes at different depths, and between various taxonomic groups of the marine microbial community? What is the relationship or is there any between community abundance and activity? To address those questions we examined the phylotype-specific activity patterns of the Eubacterial community from two depths in the open ocean using rRNA/rDNA ratio as a proxy (Chapter 3). Very recently researchers have started to use massively parallel sequencing technique to quantify the relative abundance of rRNA or rRNA/rDNA ratio as a metric of the relative activity of different phylotypes within the community (Campbell et al. 2011b; Gaidos et al. 2011; Placella et al. 2012). The strength of this method is that it provides opportunity to compare activity patterns and community structures simultaneously and it examines activity variation directly in natural environment without incubation artifacts. Our work is the first to apply this method to the open ocean ecosystem which is known dominated by microbial activities and contributes significantly to global biogeochemical cycles. However, the rRNA-activity relationship at the community level could be affected by different factors, such as taxon-specific correlation (Flynn et al. 2010), deviation from steady state (Kerkhof and Kemp 1999), and different rDNA operon numbers (Acinas et al. 2004b); not to mention known bias associated with amplicon sequencing (Berry et al. 2011; Haas et al. 2011; Pinto and Raskin 2012; Polz and
Cavanaugh 1998) and reverse transcription (Cocquet et al. 2006; Hansen et al. 2010). To overcome these limitations, more quantitative experiments sequencing independent methods are required to verify the rRNA-activity relationship for key environmental organisms.

With the tool to explore specific activity patterns in the field, the next question to answer is how environmental factors drive the activity pattern. Small size microbes with high surface to volume ratio (i.e. large exposure area to the environment) and single cell structure (i.e. direct interaction and short signaling pathway) are particularly suited for studying biological responses to environmental changes. The cellular activity variation reflects instantaneous response to the environment; later this change propagates to standing stock (biomass) distribution, and eventually it may cause long-term feedback at the evolutionary level. For *Prochlorococcus*, eHL-II and eHL-I are two high-light adapted ecotypes and temperature has been hypothesized as the major environmental factor driving their differentiation. Different biogeography patterns have been observed for these two ecotypes where eHL-II dominates lower latitude and eHL-I dominates higher latitude (Johnson et al. 2006; Zinser et al. 2007). Comparative genomic study has provided evolutionary clues of their diversification (Coleman et al. 2006; Kettler et al. 2007). Their growth rate response to temperature was examined in lab (Johnson et al. 2006) but to date there is no direct activity comparisons of these two ecotypes reported from the field. Here using two genetically and ecologically distinct
Prochlorococcus clades as a model system, we tested if temperature differences cause activity (rRNA/rDNA) differentiation between eHL-I and eHL-II in the field, and to further explored the underlying mechanism we assessed how different metabolic activities respond to temperature change in the lab (Chapter 4).

As a whole, the work presented in this thesis is aimed to reveal the taxon-specific microbial activity patterns in the field. Combined with information from biomass distribution/diversity, biotic and abiotic environmental variables along time and space, this information on in situ activity leads to a better understanding of how different phylogenetic groups contribute to ecological and biogeochemical functions differently and how microbial community as a complex system respond to environmental changes.
2. In situ activity of a dominant Prochlorococcus ecotype (eHL-II) from rRNA content and cell size

This chapter has been published as:


Authors’ contributions

ZIJ and YL conceived the study; YL, KG, AL, VPL, ERZ and JC collected field data and samples; YL undertook molecular benchwork and culture experiment; YL and ZIJ drafted the manuscript. All authors have read and approved the final manuscript.

2.1 Introduction

The single-celled cyanobacterium Prochlorococcus numerically dominates open-ocean ecosystems (Campbell et al. 1998; Campbell et al. 1997; Chisholm et al. 1988; Johnson et al. 2006; Yooseph et al. 2010; Zubkov et al. 2000b; Zvirglmaier et al. 2007) and is responsible for a substantial fraction of the total energy and carbon flux into marine food webs and biogeochemical cycles (Landry 2002; Vaulot et al. 1995). A leading
hypothesis to explain the dominance of *Prochlorococcus* across large environmental gradients is that genetically and physiologically diverse clades co-exist, but occupy different ecological niches (Ferris and Palenik 1998; Moore et al. 2002; Moore et al. 1998). Phylogenetic analyses using *rpoC1*, 16S rRNA, 16S-23S rRNA Internal Transcribed Spacer (ITS) and whole genomes have revealed distinct evolutionary lineages within this group (Kettler et al. 2007; Moore et al. 2002; Palenik and Haselkorn 1992; Rocap et al. 2002). Laboratory studies with isolates representing different genetic clades have demonstrated distinct physiological characteristics such as pigment content (Moore and Chisholm 1999; Moore et al. 1998), photophysiology (Moore and Chisholm 1999; Moore et al. 1995; Moore et al. 1998), temperature responses (Johnson et al. 2006; Zinser et al. 2007), and nutrient utilization (Martiny et al. 2006; Moore et al. 2002). Moreover, biogeographic surveys based on qPCR or dot blot hybridization techniques have revealed that these *Prochlorococcus* clades have distinct ecological distributions (Johnson et al. 2006; Scanlan et al. 1996; Zwirglmaier et al. 2007), and therefore they are considered as different ecotypes (Cohan 2006). Models that combine information from laboratory based physiology and genetic studies into dynamic global ocean environments have demonstrated that mathematical analogues of these different genetic clades have unique ecologies with implications for ecosystem structure and function (Follows and Dutkiewicz 2011; Follows et al. 2007; Rabouille et al. 2007).
However, although much is known about the genetic diversity, physiological responses and environmental distribution of some clades of *Prochlorococcus*, little is known about the *in situ* activity of these clades. Their population abundance represents the time-integrated result of *in situ* cellular growth (a type of activity), transport (both immigration and emigration) and cell losses including grazing, viral lysis or cell death. Assuming steady state, abundance is controlled by the balance between growth and losses, and patterns of activity should be related to abundance. But mechanisms such as selective grazing (Apple et al. 2011; Monger et al. 1999) and virus lysis (Sullivan et al. 2003) or departures from steady state such as nutrient pulses or water column mixing can cause an uncoupling of abundance and growth. For example, the “kill the winner” hypothesis (Thingstad 2000; Thingstad and Lignell 1997) posits that actively growing bacterial population can be efficiently removed by a virus strain, which leads to decreased abundance of the host even though it is highly active. Transient nutrient pulses from the deeper source waters (Johnson et al. 2010a) or microscale pulses from grazers (Hunt et al. 2010) can lead to transitory resources and temporal uncoupling of specific activity and abundance. Assessments of specific populations and whole communities from different environments have shown that activity and abundance are uncoupled supporting the importance of these mechanisms and in turn the necessity of quantifying both abundance and activity for understanding the mechanisms driving
microbial ecosystems (Campbell et al. 2011b; Hunt et al. 2013; Jones and Lennon 2010; Placella et al. 2012).

At the community level, the in situ specific activity of Prochlorococcus has been estimated using several techniques including size fractionated C-14 based primary production (Maranon et al. 2003; Moran et al. 2004), divinyl-chlorophyll specific pigment labeling (Goericke and Welschmeyer 1993; Redalje 1983), cell division frequency through diel flow cytometric analysis (Binder et al. 1996; Vaulot et al. 1995), and Landry-Hassett dilution experiments (Landry and Hassett 1982; Landry et al. 1995b; Worden and Binder 2003a). With numerous exceptions (e.g. (Landry et al. 2008)), these estimates of specific activity are generally coupled with abundance patterns in Prochlorococcus assemblages across environmental gradients (e.g. depth) suggesting a tightly coupled and regulated microbial community (Selph et al. 2011a). Nevertheless, these studies are limited to examining the patterns of the whole community even though its component populations are comprised of different ecotypes that are known to have different ecologies and presumably in situ activity. Thus, although there exist some estimates of microbial and phytoplankton community activity, the in situ activity of specific Prochlorococcus ecotypes remains unknown.

To address this limitation, in this study we developed a technique to estimate clade specific Prochlorococcus activity from its cellular ribosomal RNA (rRNA) 23S content detected by RT-qPCR using the dominant clade of Prochlorococcus eHL-II
(eMIT9312) as a model ecotype (Johnson et al. 2006; Zvirglmaier et al. 2007). Pioneering studies have shown that cellular rRNA correlates with cellular protein synthesis suggesting that it can be used as an index of cellular metabolic activity and potentially growth rate (Schaechter et al. 1958). For example, strong linear relationships between cellular rRNA content and growth rate have been demonstrated for *Escherichia coli* and *Salmonella typhimurium* (Bremer and Dennis 1996a; Delong et al. 1989; Dortch et al. 1983; Poulsen et al. 1993). But these relationships may be taxon-specific as there appear to be different relationships between cellular rRNA content and growth rate in cultured slow growing marine microbes (Kemp et al. 1993), cultured cyanobacteria *Synechococcus* (Binder and Liu 1998; Lepp and Schmidt 1998; Parrott and Slater 1980; Worden and Binder 2003a) and the *Prochlorococcus* strain MED4 (Worden and Binder 2003a) potentially necessitating a clade-specific approach that can target specific members of a community. Significantly, the DNA sequence of the 23S rRNA gene is sufficiently diverse to design primers to target ecologically distinct genetic clades of *Prochlorococcus* among mixed *in situ* populations of marine microbes. However, in spite of this potential, the correlation between cellular rRNA content and activity can be influenced by cell physiological characteristics such as cell size (Worden and Binder 2003a), rRNA operon copy number (Fegatella et al. 1998), and non-steady state growth (Kerkhof and Kemp 1999) as well as affected by environmental variations such as temperature (Flynn et al. 2010) and nutrient concentrations (Koch and Deppe 1971). Indeed there appears to
be dramatic variability in the rRNA/rDNA ratio among different bacterioplankton and across marine environments (Campbell et al. 2011b; Hunt et al. 2013). Recognizing that different metrics of specific activity are expected to be uncoupled in non-steady state conditions, here we seek to develop a molecular based approach to estimate in situ activity of ecologically distinct populations of a model marine microbe with global implications. We investigated (1) the quantitative relationship of cellular rRNA content (23S rRNA copy number) and cell size to specific growth rate in different strains of the eHL-II clade of Prochlorococcus to establish a relationship (if any) between cellular rRNA content and activity and (2) the patterns of cellular rRNA of eHL-II from representative depth profiles in the Pacific Ocean in relation to environmental variables and another metric of activity. We show that cellular rRNA content normalized to a metric of cell size is strongly correlated with intrinsic growth rate and that this relationship is conserved among two members of the clade. Applied to open ocean samples, this technique reveals unique insights into the ecology of this important marine microbe and provides a potential methodology for measuring the in situ activity of groups of microbial populations including those that are not currently domesticated.

2.2 Experimental Procedures
2.2.1 Culture conditions and sampling

*Prochlorococcus* strains MIT9312 and MIT9215, both members of the eHL-II (eMIT9312) clade, were grown in filtered Sargasso Sea water (0.2 µm) amended with Pro99 nutrients as described previously (Moore et al. 2002). Triplicate 25 ml batch cultures were grown at 22°C on a 12:12-h light:dark cycle with constant daytime light level and transferred to fresh media to maintain constant exponential growth. Different light levels were used to achieve different intrinsic growth rates. *In vivo* chlorophyll fluorescence was measured daily with a fluorometer (Turner 10-AU, Sunnyvale, CA). Specific growth rate was calculated by linear regression of *in vivo* fluorescence (natural log transformed) versus time (Wood et al. 2005). After acclimation for >10 generations at each of the respective light levels/growth rates, cells in middle exponential phase were harvested in the incubator morning (3 h after lights on). For nucleic acids, 1 ml of culture in replicates (2 for RNA, 2 for DNA) was filtered through a 25 mm 0.2 µm polycarbonate membrane under low vacuum (>0.3 bar). To preserve RNA, each filter was stored with 600 µl of RLT solution (Qiagen, Chatsworth, CA) with 1% β-mercaptoethanol and stored at -80°C until later extraction. To minimize any potential degradation of rRNA, the RNA sample handling time (from culture out of incubator until samples in freezer) was less than 15 min. To preserve DNA, 3 ml of preservation solution (10 mM Tris-HCl pH 8.0, 100 mM EDTA, 0.5 M NaCl)(Ahlgren et al. 2006b) was filtered through each membrane with cells and the filter was collected and stored at -
80°C until extraction. For flow cytometry analysis, 1 ml of culture was preserved with 0.125% glutaraldehyde and stored at -80°C until later analysis.

In the diel experiment, 2 L of strain MIT9312 was grown at 26°C in Pro99 media bubbled with 0.2 µm filtered air to prevent potential carbon limitation. The growth chamber (Percival Scientific, Boone, IA) was programmed to provide a 12:12-h light:dark cycle with a continuous sinusoidal irradiance pattern over sunrise and sunset to mimic the natural solar cycle. The maximum light intensity was 70 µmol Q m⁻² s⁻¹. After acclimation for more than 10 generations, cells in exponential phase (~one doubling per day) were sampled every 2 h over 26 h. DNA and flow cytometry samples were taken and preserved as previously described. Samples for two-step RT-qPCR assay were collected by freezing unpreserved culture at -80°C.

2.2.2 Field sample collection and characterization

Samples were collected from four stations – one from the Western Equatorial Pacific (2.00° N, 145.00° E) on 26 September 2006 at 13:35 (UTC time), one from the North Pacific Subtropical Gyre (12.43° N, 167.73° W) on 7 January 2007 at 19:00 (UTC time) and two adjacent stations from the North Pacific Subtropical Gyre (1) 23.91° N, 153.42° W sampled on 2 March 2012 at 15:11 (UTC time), and (2) 25.02° N, 148.89° W sampled on 3 March 2012 at 15:12 (UTC time). Hydrographic data and discrete water samples were collected from different depths using a CTD rosette system. For each depth, DNA, RNA and flow cytometry samples were taken in duplicate using the
procedure described earlier for lab cultures, except that 100 ml of seawater was filtered for DNA and RNA samples. All samples were frozen in liquid nitrogen and later transferred to -80°C freezer. Nitrate (NO$_3^-$) + Nitrite (NO$_2^-$) (N+N) from samples stored at -80°C until later analysis were quantified using an Astoria-Pacific A2 autoanalyzer following the manufacturer’s recommended protocols (fluorescence based).

2.2.3 DNA extraction and quantitative PCR

*Prochlorococcus* cell abundance was measured by a quantitative PCR (qPCR) assay as described in (Johnson et al. 2006) using eHL-II specific 23S rRNA primers (Ahlgren et al. 2006b).

2.2.4 RNA extraction and RT-qPCR

Based on the specifics of sampling and availability of kits/reagents, three different procedures (A – C) were used to extract RNA and perform RT-qPCR. To compare results between different procedures a control sample (replicates preserved from a single culture) was included in each batch and all RT-qPCR results were normalized to procedure C which yield the highest cDNA per cell. (A) For the culture light-limit experiment (Fig. 2) and the two field profiles used to compare to photosynthesis (Fig. 4), cells collected on filters were lysed by bead-beating (BioSpec, Bartlesville, OK) with 0.1 mm Zirconium beads (BioSpec) at 4,800 rpm for 2 min with cooling on ice for every 30 sec. Total RNA was extracted from the lysate using RNeasy Mini kit (Qiagen) with DNase I on column digestion according to the manufacturer’s
instructions. cDNA was synthesized from extracted RNA using iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA) as described in the manufacture’s instruction. Synthesized cDNA was diluted 1:100 in RNase-free water and used as the template for qPCR. (B) For the diel experiment, two-step RT-qPCR was performed using Power Syber Green Cell-to-Ct kit (Ambion, Austin, TX) as described in the manufacture’s protocol, with two modifications. First, in the cell lysis step 5 µl of thawed cell culture was added to 50 µl of Lysis Solution containing 1% DNase I, and the mixture was incubated at room temperature for 8 min. Second, the synthesized cDNA was diluted 1:10 in RNase-free water and used as template in following qPCR reaction. (C) For the two adjacent field profiles (Fig. 5), cells were lysed as described in procedure A. Then the lysate was diluted 1:100 in RNase-free water and used directly in cDNA synthesis (iScript cDNA Synthesis Kit). Because no DNase digestion step, gDNA copies determined by qPCR were subtracted from the RT-qPCR results.

2.2.5 Flow cytometry

*Prochlorococcus* cells were enumerated using a FACSCalibur flow cytometer (Becton Dickinson) and populations characterized as previously described (Johnson et al. 2010b). Briefly, cells were excited with a 488 nm laser (15 mW Ar) and inelastic forward (<15°) scatter, inelastic side (90°) scatter (SSC), green (530 ± 30 nm) fluorescence, orange fluorescence (585 ± 42 nm), and red fluorescence (> 670 nm) emissions were measured. Population mean properties (scatter and fluorescence) were normalized to
1.0 μm yellow green polystyrene beads (Polysciences, Warrington, PA). Heterotrophic bacterioplankton were quantified by staining the samples with SYBR Green –I (Molecular Probes Inc.) (Marie et al. 1997) and subtracting Prochlorococcus from the total bacterioplankton population.

2.2.6 Photosynthesis (Primary production)

Photosynthesis (primary production) was estimated from radiolabeled carbon tracer incorporation during 24-h on-deck incubations as previously described (Barber et al. 1996; Lance et al. 2007). Briefly, water samples were collected into 60 mL polycarbonate bottles and inoculated with 10 μCi of 14C bicarbonate solution. For each station, one replicate surface sample was filtered immediately to provide time-zero controls. The remaining samples were incubated for 24 h on deck in acrylic incubators screened with blue plus neutral density stage screening to achieve a given percentage of surface irradiance, nominally 100%, 47%, 30%, 16%, 10% and 1%. Depth of each light level was calculated using an inverse analytical radiative transfer model as previous described (Johnson et al. 2002). Incubator temperatures were maintained by a continuous flow of surface seawater. After incubation, duplicate samples were filtered onto 0.2μm polycarbonate or Whatman GF/F filters then acidified with 0.5 mL of 0.5 N HCl for 24 h to liberate unincorporated inorganic 14C. Ecolume scintillation fluid (7 mL) was added and activity quantified. Total added 14C, was quantified in select inoculated subsamples prior to filtering. Primary production was estimated from the activity of the
total and filtered samples using standard calculations (Barber et al. 1996). When an exact depth match was not present, primary production values were linearly-interpolated to the depth of molecular samples using nearest neighbor values.

2.3 Results and discussion

2.3.1 Relationship of rRNA and cell size to specific growth rate

To determine the relationship of rRNA and cell size with specific growth rate, we stimulated different intrinsic growth rates by regulating light intensity, which is the most significant environmental variable structuring the Prochlorococcus community and distributions in the field (Johnson et al. 2006; Martiny et al. 2009a; Zinser et al. 2007) (Fig. 2A). At low light levels and below a threshold irradiance (Bittar et al. in review; Moore et al. 1995), the growth rates for both strains increase rapidly with additional light and, although the eHL-II clade is considered a high-light ecotype, growth saturates near ~40 µmol Q m\(^2\) s\(^{-1}\) (Moore et al. 1995). The maximum growth rate obtained was 0.71 d\(^{-1}\) for MIT9312 and 0.53 d\(^{-1}\) for MIT9215, which is slightly lower than previously reported (Moore and Chisholm 1999), but nevertheless consistent with the lower growth temperature used in this study (22°C vs. 24°C).
According to the Mie solution to Maxwell’s equations, both forward angle scatter (FSC) and side scatter (SSC) vary as a function of cell diameter (and volume) and refractive index, which itself is a function of cellular biochemical composition including carbon content (Ackleson and Spinrad 1988). Empirical relationships have been established between cell volume and FSC (Cavender-Bares et al. 2001; Simon et al. 1994) or SSC (Felip et al. 2007; Simon et al. 1994) for a variety of phytoplankton groups and they generally follow Mie’s solutions even though the differential contribution of cell volume and cellular composition (e.g. carbon quota) to FSC or SSC is difficult to experimentally quantify, but is theoretically different for FSC and SSC respectively (Ackleson and Spinrad 1988). Here we use SSC as an operational index of cell size (which contains both cellular composition and volume components) because experimentally it can be more precisely quantified and because Mie’s solution predicts that it is more substantially influenced by cell composition (i.e. carbon quota) than volume per se (Ackleson and Spinrad 1988), which we expect to influence cellular activity more than the volume per se. Comparisons using FSC resulted in similar, though less robust, relationships.
Table 2: Statistical parameters of the linear regressions shown in Fig. 3B-F

<table>
<thead>
<tr>
<th>Strain</th>
<th>Variable</th>
<th>Two-way least square fit</th>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>x y Slope (SD) y-intercept (SD) r² P-value</td>
<td></td>
</tr>
<tr>
<td>MIT9215</td>
<td>µ RNA cell⁻¹</td>
<td>150 (33) 150 (12) 0.87 0.023</td>
<td></td>
</tr>
<tr>
<td>MIT9215</td>
<td>µ SSC</td>
<td>-0.055 (.009) 0.043 (.003) 0.93 0.008</td>
<td></td>
</tr>
<tr>
<td>MIT9312</td>
<td>µ RNA cell⁻¹</td>
<td>400 (130) 170 (67) 0.71 0.042</td>
<td></td>
</tr>
<tr>
<td>MIT9312</td>
<td>µ SSC</td>
<td>-0.063 (.008) 0.066 (.004) 0.93 0.002</td>
<td></td>
</tr>
<tr>
<td>MIT9215, MIT9312</td>
<td>µ RNA cell⁻¹ SSC⁻¹</td>
<td>28000 (2700) -700 (1200) 0.92 &lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: (A) Specific growth rate ($\mu$) as a function of irradiance and (B) side scatter (SSC) as index of cell size, (C) rRNA cell$^{-1}$, and (D) rRNA cell$^{-1}$ normalized by SSC for MIT9312 (open circles) and MIT9215 (closed circles) as a function of growth rate. Mean rRNA cell$^{-1}$ SSC$^{-1}$ vs. growth rate of the diel experiment is shown as gray square in (D).

The cell sizes (SSC) of both MIT9312 and MIT9215 are anticorrelated with light levels (and specific growth rate) (Fig. 2B; Table 2), which is the opposite of the trend observed in many model heterotrophic bacteria; faster growing cells typically have more DNA, RNA and protein per cell and therefore are larger (Bremer and Dennis 1996a; Schaechter et al. 1958; Shiomi and Margolin 2007). This difference between autotrophic and heterotrophic cells may be the result of packing of additional pigments and other
components of the photosynthetic apparatus (thus larger size) at lower irradiance
(Worden and Binder 2003a), a general trend has been observed by others (Cavender-
Bares et al. 2001). Regardless, the cell size of strain MIT9312 is significantly different
from strain MIT9215 throughout the range of growth rates (Fig. 2B), perhaps reflective of
their different isolation environments (135 m Gulf Stream Atlantic Ocean for MIT9312
and surface Equatorial Pacific for MIT9215 (Moore and Chisholm 1999)).

rRNA cell$^{-1}$ and specific growth rate are linearly correlated for both strains
(MIT9215, $R^2 = 0.87$ and MIT9312, $R^2 = 0.71$) across the entire range of growth rates (Fig.
2C). This pattern is different from a previously reported relationship for Prochlorococcus
strain MED4 (a member of clade eHL-I) in which rRNA cell$^{-1}$ increases at intermediate
growth rates and decreases near the maximum growth rate (Worden and Binder 2003a),
which may be reflective of different clade responses. The slope of the MIT9312 linear fit
is significantly greater ($P<0.01$) than MIT9215 and is likely due to the larger size of
MIT9312 (Fig. 2B). Assuming a constant ribosomal efficiency, larger cells with higher
protein content need to synthesize additional protein and therefore require additional
rRNA per cell to maintain this additional demand (Schaechter et al. 1958).

To account for this variability with size, we normalized rRNA cell$^{-1}$ by SSC and
found that both MIT9312 and MIT9215 fit to a single linear relationship between rRNA
cell$^{-1}$ SSC$^{-1}$ and growth rate (Fig. 2D) and the correlation is improved ($R^2 = 0.92$). Because
the slopes of the linear fits to each individual strain are the same as that of both strains

34
together (p>0.1), we hypothesize that there may exist a conserved relationship between rRNA cell\(^{-1}\) SSC\(^{-1}\) and growth rate for this *Prochlorococcus* ecotype (eHL-II), which may extend to other clades of *Prochlorococcus* or potentially other marine microbes. For groups where this relationship is conserved, rRNA cell\(^{-1}\) SSC\(^{-1}\) could potentially be a powerful tool to assess the specific activity including specific growth rate of natural *Prochlorococcus* populations within mixed communities and including uncultivated genotypes.

### 2.3.2 Diel variation

*Prochlorococcus* cell size, DNA content, photophysiology and gene expression are tightly coupled with light intensity over the day and night cycle (Brunet et al. 2008; Bruyant et al. 2005; Zinser et al. 2009). These properties (and others) are tightly choreographed to optimize growth to predictably variable light, but may influence the instantaneous relationship between cellular RNA content and growth rate potentially limiting this approach. To investigate this possibility, MIT9312 was grown under a simulated daily sinusoidal light-dark cycle and was sampled every 2 h over a 26 h period. The daily specific growth rate was 0.73 d\(^{-1}\), which is approximately one doubling per day (Fig. 3A) such that the division cycle of the entire population was synchronized to maximize any diurnal variability in cellular composition and therefore population characteristics.
Figure 3: Diel variation in *Prochlorococcus* strain MIT9312 of (A) cell concentration, (B) side scatter (SSC), (C) total population 23S rRNA, (D) rRNA cell$^{-1}$ and (E) rRNA cell$^{-1}$ SSC$^{-1}$ (closed triangles) and relative irradiance (solid line). Shaded area represents incubator dark period.
Cell division occurred at night with total cell counts increasing rapidly after sunset and becoming stable before sunrise (Fig. 3A) and is consistent with previous studies for other strains (Bruyant et al. 2005; Vaulot et al. 1995; Zinser et al. 2009). During the daytime cell size increased by ~2-fold as a result of biomass accumulation from photosynthesis followed by decreases in size at night from cell division (Fig. 3B). Overall, these patterns are consistent with those previously reported for Prochlorococcus strain MED4 (Zinser et al. 2009) and show an organism tightly coupled with the light cycle. Similarly, total population rRNA was synthesized and accumulated during the daytime, but remained relatively constant at night, even as cells were dividing (Fig. 3C). Taken together, the cellular rRNA expression level (rRNA cell$^{-1}$) has a substantial diel cycle (CV = 25%, Fig. 3D). This expression pattern is driven by cells accumulating biomass (photosynthate) during the daytime requiring more ribosomes for total protein synthesis, and the energy demand for ribosome synthesis can be provided via active photosynthesis at this time. However, both rRNA cell$^{-1}$ and SSC show dramatic diel patterns that are similar in magnitude (and the mechanisms driving them are shared) so when rRNA cell$^{-1}$ is normalized to SSC, rRNA cell$^{-1}$ SSC$^{-1}$ has a reduced diel variation (CV = 15%, Fig. 3E). Indeed, the variation associated with the diel cycle is comparable to the error range of this technique (e.g. the standard deviation between biological replicates sampled from the same time point) (Fig. 3D). Thus, while sampling should occur at the same time of day to minimize other biases associated with diel variability,
diel effects on rRNA cell$^{-1}$ SSC$^{-1}$ are small compared to other sources of variability (i.e. changes in growth rate associated with environmental variables), therefore sampling time of day is not critical for this estimation.

2.3.3 Field patterns and relationship to environmental variables

To examine the suitability of this technique for mixed population field samples, we explored the vertical variation of eHL-II activity as assessed by rRNA and cell size from multiple stations in the Pacific Ocean. Initially, we tested this approach by comparing total eHL-II rRNA transcript number to community photosynthesis from two stations in the Pacific Ocean, including one from the North Pacific Subtropical Gyre (Fig. 5A, station #1) and the other from the Western Equatorial Pacific (Fig. 5B, station #2), where *Prochlorococcus* dominates the phytoplankton community, where eHL-II dominates the *Prochlorococcus* community (>70%) and where *Prochlorococcus* contributes substantially to the total net primary production (Vaulot et al. 1995). Unlike rRNA cell$^{-1}$ SSC$^{-1}$, which is a measure of cell-specific activity, total eHL-II rRNA is proportional to eHL-II community activity and therefore related to biomass production. Although the maximum water column primary production ($P_{\text{opt}}$) varies by ~10-fold between the two locations due to differences in the availability of nutrients and the levels of phytoplankton biomass (data not shown), there are corresponding differences in the total eHL-II rRNA. This relationship ($R^2=0.85$) between total eHL-II rRNA and
photosynthesis (carbon uptake) (Fig. 4C) further supports the relationship between rRNA and cellular activity, here quantified by photosynthesis.

Figure 4: Vertical profiles of $^{14}$C based primary production (open symbols) and total eHL-II rRNA (closed symbols) for stations in (A) the North Pacific Subtropical Gyre (NPSG) (Station #1, circles) and (B) the Western Equatorial Pacific (Station #2, squares) and (C) The relationship between rRNA and photosynthesis (carbon uptake) (circles – Station #1, squares – Station #2).

To further explore the utility of this approach and the variance in specific activity in a natural setting, we sampled two geographically close, but oceanographically distinct stations from North Pacific Subtropical Gyre (NPSG). Samples were collected within the euphotic zone from 0 to 200 m where Prochlorococcus numerically dominates the
phytoplankton community. The water column at the first site (station #3, Fig. 5A) was moderately mixed with a mixed layer depth (MLD) of ~94 m (as inferred from temperature profiles) and a nitracline at ~75 - 100 m (defined by the depth \([\text{NO}_2^- + \text{NO}_3^-] > 100 \text{ nmol l}^{-1}\) (Winn et al. 1995) whereas at the second station (station #4, Fig. 5B) there was a deep vertically mixed layer (~175 m) and a much deeper nitracline at ~165 - 200 m. While both stations are oligotrophic and ideal habitats for the \textit{Prochlorococcus} community, each represents a unique combination of environmental variables with potentially unique activity signatures.

At both sites, eHL-II was the most abundant clade of \textit{Prochlorococcus} (>70% of the total \textit{Prochlorococcus} abundance) and roughly constant in the upper mixed layer (Figs. 5B and F) with similar maximum cell concentrations ~10^5 cells ml\(^{-1}\). However, unlike heterotrophic bacterioplankton, the abundance of \textit{Prochlorococcus} eHL-II decreased rapidly below the mixed layer at both sites, consistent with prior field observations and also consistent with their designation as a ‘high-light’ ecotype (Ahlgren et al. 2006b; Johnson et al. 2006; Malmstrom et al. 2010; West and Scanlan 1999). Similarly, at both stations \textit{Prochlorococcus} cell size is generally constant and small within the mixed layer (Figs. 5C, G), then increases with depth below the mixed layer due to photoacclimation and increased pigment content (Fig 5C). This pattern is also consistent with the relationship observed in the laboratory between cell size and specific growth rate as a function of light-limitation (Fig. 2). Throughout both of the water columns, there were
no discernible flow cytometrically defined subpopulations of Prochlorococcus (i.e. cell size and pigmentation were homogenous within the community at given depth).

To explore the ecotype specific activity, we estimated eHL-II rRNA content at these two sites using RT-qPCR. We assumed that the cell size of the eHL-II subpopulation is the same as the mean size of the total Prochlorococcus community because there are no significant subpopulations evident in the SSC histograms and the Prochlorococcus community is dominated by eHL-II. We normalized eHL-II rRNA content by cell numbers estimated from eHL-II qPCR (Johnson et al. 2006) and cell size (SSC), then converted to estimated growth rates (d⁻¹) using the relationship found in controlled laboratory experiments (Fig. 2, Table 2). At both sites in the upper 50 m, the estimated eHL-II specific growth rates ranged from 0.2 to 0.4 d⁻¹. At the moderate mixing site (station #3), the estimated eHL-II maximum growth rate within the water column is ~ 0.4 d⁻¹, which is comparable to the maximum Prochlorococcus growth rates (0.1 to 0.5 d⁻¹) estimated in oligotrophic waters through ¹⁴C labeled divinyl-chlorophyll a (Goericke and Welschmeyer 1993). At the deep mixing site (station #4) the estimated eHL-II maximum growth rate was ~ 1.0 d⁻¹ which is more similar to the maximum growth rates estimated in the larger Equatorial Pacific Ocean from cell division frequency analysis (0.6 to 1.4 d⁻¹) (Mann and Chisholm 2000; Vaulot et al. 1995).

Although estimates from both stations are generally consistent with the few available prior studies, both specific activity and inferred growth rates vary
substantially within each water column and do not follow the same trends with depth as cell concentrations. For example, at the moderate mixing site (station #3) eHL-II specific activity \((\text{rRNA cell}^{-1} \text{SSC}^{-1})\) has a subsurface maximum at ~50 m and then decreases with depth (Fig. 5D), suggesting photoinhibition near the surface and subsequent light-limitation below 50 m. This light dependent rRNA-based specific activity profile is consistent with \textit{Prochlorococcus} cell division frequency profile previously reported in the equatorial Pacific (Vaulot et al. 1995). Conversely, at the deep mixing site, rRNA cell\(^{-1}\) SSC\(^{-1}\) increases from the surface to the base of the mixed layer to a maximum at 125 m (Fig. 5H), beyond which it decreases rapidly. This deep subsurface peak in rRNA cell\(^{-1}\) SSC\(^{-1}\) is above the nitracline (Fig. 5E) and may be the result of N flux stimulating the growth of the eHL-II population. Although eHL-II is a nominal “high-light” ecotype, laboratory-measured growth saturation irradiance for eHL-II isolates are low (<40 µmol Q m\(^{-2}\)s\(^{-1}\)) (Moore et al. 1999, this study) corresponding to deeper within the water column. Previous studies have also shown enhanced \textit{Prochlorococcus} specific activity near the nutricline and well below the surface layers (Hunt et al. 2013; Johnson et al. 1999; Vaulot et al. 1995) suggesting that low light levels are adequate to maintain elevated specific activity levels. Specific activity at depth and near the chlorophyll maximum may also be enhanced by heterotrophic mechanisms (Zubkov et al. 2003) that may be particularly important in higher specific grazing (and dissolved organic flux) environments (Barber and Hiscock 2006; Landry et al. 2011). Regardless of the precise
mechanisms that establish the location of the maxima in specific activity, these vertical profiles suggest that in these open ocean environments eHL-II specific activity (rRNA cell$^{-1}$ SSC$^{-1}$) is driven by the interplay of light and nutrient availability ($\text{NO}_3 + \text{NO}_2$) and likely other mechanisms.

Surprisingly, in spite of significant vertical variability in eHL-II specific activity, both stations have relatively constant cell concentrations within the upper mixed layer, which decrease dramatically below the thermocline (Fig. 5). Given the vertical variability in eHL-II specific activity (rRNA cell$^{-1}$ SSC$^{-1}$), but relative constant biomass patterns with depth within the mixed layer, vertical mixing, through mechanisms such as diurnal convective water column overturning or wind-driven mixing, may be redistributing cells in the water column to balance any differences in depth-specific activity, but not to the degree that specific activity is uniform within the water column. Advection of narrow vertical bands or even thin layers may also introduce differences in apparent local activity and abundance (Durham and Stocker 2011). Alternatively (and not mutually exclusive), there may be a decoupling between specific activity and biomass profiles due to differential loss pressures such as viral lysis or grazing that specifically targets highly active or fast growing cells (Apple et al. 2011; Thingstad and Lignell 1997). This apparent decoupling of specific activity and abundance in these open ocean environments (Campbell et al. 2011b; Hunt et al. 2013; Landry et al. 2011) has potential implications for ecosystem structure and function.
Figure 5: Vertical profiles of (A, E) light (% surface irradiance), temperature (°C) and nitrate + nitrite (N+N, µM), (B, F) eHL-II and heterotrophic bacterioplankton abundance (cells ml⁻¹), (C, G) side scatter (SSC) of *Prochlorococcus* relative to 1.0 µm polystyrene beads and (D, H) specific activity of eHL-II (rRNA cell⁻¹ SSC⁻¹) for representative locations in a moderate mixing site (A-D) and a deep mixing site (E-H) in the North Pacific Subtropical Gyre.
2.4 Conclusion

Our study demonstrates that MIT9312 and MIT9215, two strains of the most globally abundant clade of *Prochlorococcus* (eHL-II), have robust linear relationships between rRNA cell$^{-1}$ and specific growth rate as well as between cell size (SSC) and specific growth rate, although the slopes for each strain are different. After cellular rRNA content is normalized by cell size, both strains show the same overall relationship suggesting a conserved correlation between rRNA cell$^{-1}$ SSC$^{-1}$ and specific growth rate for this ecotype (and potentially others) and supporting the potential of using rRNA cell$^{-1}$ SSC$^{-1}$ as an index of *in situ* specific activity. Applying this method at two stations with different oceanographic conditions, rRNA cell$^{-1}$ SSC$^{-1}$ of eHL-II is variable with depth and strongly influenced by environmental variables, whereas abundance alone does not show this variability. These patterns provide additional insight into the ecology of these populations, including the balance between growth and losses or physical transport, and suggest mechanisms regulating microbial populations and ultimately their contribution to ecology and biogeochemistry. Taken together, these results demonstrate the utility of cell size and rRNA content in uncovering clade-specific patterns of marine microbial activity. Using rRNA as an index of *in situ* activity in combination with water mass movement or population specific *in situ* loss terms (e.g. grazing, lysis, etc.) over space and time could be powerful tools to study the biological and physical mechanisms in
regulating microbial populations in a dynamic and changing global ocean. Future studies should also address how this specific activity metric may apply to other clades of marine microbes as well as the relationship between rRNA and other forms of cellular activity including how this relationship varies in response to other environmental variables such as nutrients, temperature or non-steady state conditions.

2.5 Acknowledgements

This work was supported in part by NSF # 0959630 and 1031064 to ZIJ and NSF #1030518 to ERZ. The authors thank Maria de Oca for providing flow cytometry data. The authors declare no competing financial interests.
3. The relationship between abundance and specific activity of bacterioplankton in open ocean surface waters

This chapter has been published as:


Authors’ contributions

ZIJ, YL, and DEH conceived the study; YL collected samples; YL, DEH, SGT, ZIJ and LKI undertook molecular benchwork and analyses; DEH, YL, and ZIJ drafted the manuscript. All authors have read and approved the final manuscript.

3.1 Introduction

The application of massively parallel DNA sequencing to marine microbial ecology has increasingly identified the breadth of bacterioplankton phylogenetic and genomic diversity (Acinas et al. 2004a; Coleman and Chisholm 2011; Delong et al. 2006).
In general, the most abundant environmental sequence types have been assumed to be the most active, both on a per cell basis and as members of the community, and thus they are assumed to drive nutrient and energy cycling, although there are some important exceptions (e.g. N\(_2\)-fixers). Robust tools that can determine the activity of specific phylotypes in the environment have not been widely applied even though the ability to measure \textit{in situ} activity will address key questions in microbial ecology and biogeochemistry including the environmental variables that favor specific taxa (Johnson et al. 2006), the role of predation in regulating bacterioplankton abundance (Apple et al. 2011a), and the relative importance of specific taxa in biogeochemical cycling (Strom 2008) among others.

Culture and genome-based studies of a few model organisms often partition marine microbes into oligotrophs (e.g. \textit{Prochlorococcus} and \textit{Pelagibacter}) that grow slowly on low levels of nutrients and copiotrophs (e.g. \textit{Roseobacter}, \textit{Alteromonas}) that require high concentrations of nutrients and may display boom or bust populations dynamics (Polz et al. 2006). Although clearly delineated oligotrophs and copiotrophs are useful as a binary classification, reality may reflect a gradient in responsiveness to environmental conditions (Lauro et al. 2009). Some copiotrophs may cycle between rare and abundant states while others remain rare in spite of high apparent growth rates due to predation pressure (Campbell et al. 2011) and oligotrophs may be more dynamic than previously thought (Polz et al. 2006). Thus, combining measures of specific activity and abundance
can provide a more complete assessment of bacterioplankton community structure and its members’ impacts on biogeochemical cycling and microbial ecology.

Under steady state growth and in the absence of selective predation by viruses or grazers, the most active bacteria would be the most abundant, with higher growth rates leading to more biomass (Pedrós-Alió 2012). Based on these assumptions, rare bacteria have often been assumed to be composed of slow-growing or dormant bacteria that act as a “seed bank” becoming abundant when positively selected for by the environment (Lennon and Jones 2011). Yet there is increasing evidence that rare bacteria may be disproportionately active relative to their abundance (Campbell et al. 2011; Jones and Lennon 2010). Observations of uncoupled specific activity and abundance in a coastal ocean time series and in lakes suggest that apparent equilibrium in these environments masks a dynamic system, and that some low abundance taxa may contribute disproportionately to ecological and biogeochemical processes relative to their abundance (Campbell et al. 2011; Jones and Lennon 2010).

Activity of marine microbes has historically been assessed at the community level by measuring the incorporation of labeled precursors (Fuhrman and Azam 1982; Kirchman et al. 1985b). However, this technique cannot determine the activity of specific populations, which may differ from that of the total community. Other phyla-specific measurements often require incubations that may distort microbial populations due to “bottle effects,” incorporation of substrates or development of
specific probes/primers. To address this gap, there is a growing trend of obtaining phylotype-associated specific activity measurements by quantifying ribosomal RNA (rRNA) and ribosomal RNA genes (rDNA) that reflect the activity and abundance of specific ribotypes, respectively (Campbell et al. 2011; Jones and Lennon 2010). Cellular ribosome abundance has been shown to correlate well with growth rates of marine microbes in culture (Kemp et al. 1993; Worden and Binder 2003c) (with some noted exceptions (Kerkhof and Kemp 1999; Morgenroth et al. 2000)) and without a requirement for incubation or substrate incorporation like many other approaches. However, there are limitations to using this technique to infer an in situ growth rate as the rRNA content per cell is dependent on cell size, growth rate, a taxon-specific relationship with growth, and likely other clade-specific and environmental variables (Kerkhof and Kemp 1999). While the abundance of rDNA per cell (i.e. rDNA operons per cell) can vary by roughly one order of magnitude between taxa (Acinas et al. 2004b), the number of copies of rRNA per cell has an even wider range (0 in a dead cell to >10,000 copies/ cell (Fegatella et al. 1998)). Further, in rRNA and rDNA library based approaches, the number of reads for each taxon observed is relative to reads for other organisms in that library rather than an absolute measure of abundance. Many comparisons of rRNA and rDNA have been based on relatively small libraries that are potentially biased in their interpretation because of insufficient sampling depth. These
considerations complicate direct comparisons between rDNA and rRNA in library sequencing (Gihring et al. 2012).

Taking advantage of the strengths of this approach, while acknowledging that this technique, like others, has limitations, here we apply the ratio of rRNA to rDNA for a given ribotype as a proxy for bacterial specific activity, which likely reflects both rates of biomass and non-biomass producing cellular processes. The strength of measuring rRNA/rDNA ratios comes not from single time point but in comparing measurements across time and space to identify the drivers of activity and abundance for these organisms at representative locations in the ocean. In contrast to a previous study that examined specific activity using a coastal time series (Campbell et al. 2011), we examine open ocean stations and variability in specific activity with depth. We focus on determining specific activity in the cyanobacterium Prochlorococcus which is important both as a model organism and as the dominant clade in the tropical and subtropical open ocean. As an organism with well-known nutrient requirements and light driven physiology, measuring Prochlorococcus specific activity will allow us to identify potential uncoupling of abundance and specific activity in the environment. Overall, this study allows us to examine the specific activity of different populations of bacteria across their broad genetic diversity to better understand their respective roles in biologically driven ocean processes.
3.2 Materials and Methods

3.2.1 Sample Collection

Seawater for 16S rRNA and rDNA library construction was collected during HOT cruise 215 at Station ALOHA (22°45′N, 158°W), at two depths (25 and 100 m), on September 25, 2009, 2215 UTC. Bacterioplankton samples for DNA and RNA extraction were collected as described previously (Delong et al. 2006; Frias-Lopez et al. 2008) with slight modifications. Briefly, seawater was pre-filtered through a 125 mm Whatman GF/A filter to remove large eukaryotes. For DNA extraction, each sample was collected onto a 0.22 µm Sterivex filter cartridge (Millipore), covered with 2 ml of preservation buffer (50 mM Tris pH 8.3, 40 mM EDTA, and 0.75 M sucrose) and stored at -80°C. In total 11 and 19 L of seawater were filtered for DNA samples from 25 m and 100 m, respectively. Material for RNA extraction was collected by duplicate filtration of 1 L of seawater from the same sample used for DNA collection on 0.22 µm Durapore filters (Millipore). After filtration, each filter was immediately immersed in 1 ml of RNALater (Ambion) and frozen at -80°C until extraction. Environmental variables were obtained from the Hawaii Ocean Time-series (HOT) website (http://hahana.soest.hawaii.edu/hot/hot-dogs).

Additional depth profiles for Prochlorococcus qPCR and RT-qPCR were taken at a station in the Equatorial Pacific (2°S, 155°W), on Sep 2, 2006, 0207 UTC, and a station in the Sargasso Sea (30°43′N, 72°41′W), on May 24, 2010, 1832 UTC. At each station discrete
water samples at ~ 10 depths in the upper ~ 200 m were taken using a CTD rosette system. At the Equatorial Pacific site, macronutrient samples were collected in bottles, frozen and then analyzed on shore using an Astoria Autoanalyzer as described previously (Johnson et al. 2010b). The detection limit for nitrate is 0.05 µmol L\(^{-1}\). At the Sargasso Sea site, nitrate was measured using an In Situ Ultraviolet Spectrophotometer (ISUS) nitrate sensor (Satlantic) mounted on the CTD rosette (Johnson 2010; Johnson and Coletti 2002). The ISUS nitrate detection limit is 0.5 µmol L\(^{-1}\), and values below the detection limit were reported as zero. To obtain material for qPCR, 100 ml of seawater was filtered through a 25 mm 0.22 µm polycarbonate filter under low vacuum (>0.3 bar), washed with 3 ml of preservation solution (10 mM Tris, 100 mM EDTA, 0.5 M NaCl, pH = 8.0), and frozen at -80°C until extraction. In obtaining samples for Prochlorococcus RT-qPCR, 100 ml samples were taken in duplicate, and each filter was preserved in 600 µl of RLT solution (Qiagen) with 1% β-mercaptoethanol and frozen at -80°C until extraction.

### 3.2.2 Nucleic acid extraction for 16S rRNA and rDNA libraries

The preservation solution in the Sterivex cartridge was collected and concentrated using an Amicon Ultra 10 kDa filter unit (Millipore) and combined with the filter. DNA extraction followed the manufacturer’s instructions (DNeasy Tissue Kit, Qiagen) with the addition of bead beating (0.2 g 0.1 mm Zr beads for 3 min at 4,800 rpm). Total RNA was extracted as described previously (Frias-Lopez et al. 2008) with some
modifications. Briefly, the RNAlater solution was concentrated to less than 100 µl using an Amicon Ultra 10 kDa filter unit, then returned to the vial with filter. 0.2 g of 0.1 mm Zr beads was added to the vial, and the sample was lysed by bead-beating at 4,800 rpm for 2 min at 4°C. Total RNA was isolated from lysate using RNeasy MinElute column (Qiagen) following manufacturer’s instructions, and treated with DNase using Turbo DNA-free kit (Ambion). cDNA was synthesized using SuperScript II reverse transcriptase system (Invitrogen) with random hexamer primers.

3.2.3 Amplicon Library Preparation

16S gene fragments were amplified by PCR from environmental DNA and cDNA templates using universal primers 926F and 1392R plus barcodes and adaptors for 454 pyrosequencing. Triplicate PCR reactions (20 µl) for each library included: 0.25 Units JumpStart Taq (Sigma), 1x JumpStart PCR buffer, 0.05 µM of each primer, 200 µM of each dNTP, and 300 µg/ml BSA. The PCR reaction was thermocycled: 95°C for 3 min, 25 cycles at 95°C for 30 s, 55°C for 45 s, and 72°C for 90 s followed by a final extension of 10 min at 72°C. PCR products were separated by electrophoresis on TAE agarose gels and purified using MinElute Gel Extraction Kit (Qiagen). Final 454 library preparation and sequencing was conducted at the Joint Genome Institute (JGI) in Walnut Creek, CA.

3.2.4 Analysis of 16S rRNA and rDNA tag libraries

A total of 1.6 million raw sequences were obtained by sequencing PCR amplicons for the four libraries described above. Sequences were processed using
QIIME 1.2.1 (SVN1920) (Caporaso et al. 2010). Briefly, sequences were filtered for length (300≥ length ≤540), quality score (mean>30), and homopolymer runs <8. Sequencing errors were minimized using Denoiser 0.91 for Titanium reads, chimera removal using ChimeraSlayer, and by defining operational taxonomic units (OTUs) at 97% identity using the uclust algorithm. OTUs were removed with only one occurrence in all four libraries, with a Eukaryotic or Archaeal best BLAST hit (refseq_genomic) or an E value >10^{-30} against the SILVA (v106) database (Pruesse et al. 2007). This analysis yielded a total of 861,659 sequences after pre-processing; to control for variation in the number of reads per library the libraries were each subsampled 75,000 times (Gihring et al. 2012). For the 50 most abundant OTUs, a representative sequence was hand aligned using ARB (Ludwig et al. 2004) and a maximum likelihood tree constructed using PhyML (default parameters) with 100 bootstrap replicates (Guindon and Gascuel 2003) and a lane mask.

### 3.2.5 Prochlorococcus Ecotype qPCR and RT-qPCR DNA extraction

Filters for qPCR were shaken without beads in a bead beater (BioSpec) for 2 min at 4800 rpm with 650 µl of 10 mM Tris (pH 8.0) buffer, followed by incubation at 95°C for 15 min. The resulting lysate was used as template for qPCR reactions in parallel with serially-diluted standards of known cell concentrations. For RT-qPCR, individual samples were lysed with 0.2 g of zirconium beads in the bead beater at 4800 rpm alternating bead beating (30sec) and cooling on ice for a total of 2 minutes of bead beating. Total RNA was extracted from the sample using RNeasy Mini column (Qiagen)
with DNase digestion according to the manufacture’s protocol. cDNA was synthesized using random hexamer primers from the extracted RNA using iScript cDNA Synthesis Kit (Bio-Rad), and used as template in qPCR reactions following the same protocol as genomic DNA qPCR.

3.2.6 *Prochlorococcus* Ecotype specific qPCR and RT-qPCR

For environmental DNA samples, qPCR was performed using primers targeting the 23S rRNA designed for specific ecotypes. Ecotype MIT9312 (eHL-II) was amplified as previously described (Ahlgren et al. 2006; Johnson et al. 2010b) and ecotype MIT9313 (eLL-IV) was amplified using modified primers 23S-678F (5’-CGAGTCTGAATAGGGCGATC – 3’) and 23S-1210R (5’–CTCCCCTACCATTAAACAAG  – 3’) and thermocycled: 95°C for 15 min, 45 cycles at 95°C for 15 s, 62°C for 30 s, and 72°C for 30 s followed by a final extension of 5 min at 72°C. The specificity of the eMIT9313 primers was verified using lab isolates of *Prochlorococcus* and *Synechococcus* as well as environmental clone libraries. For synthesized cDNA, qPCR was performed using 23S rRNA gene primers for ecotypes eMIT9312 (eHL-II) (Ahlgren et al. 2006) and eMIT9313 (eLL-IV) (23S-678F and 23S-1210R). The relative abundance and 23S rRNA gene expression level of *Prochlorococcus* ecotypes from environmental DNA and cDNA pools, respectively, was determined by comparing the threshold cycle (C_T) value with serially-diluted culture standards. Because the template preparation process for the standards (Ahlgren et al. 2006) differs
from that of the environmental DNA and cDNA samples, the qPCR results are presented as unitless relative values.

Sequences are deposited in the Genbank Short Read Archive under accession number SRA058462. Oceanographic data is available at http://bcodmo.org.

Supplemental Tables are available at http://aem.asm.org/content/79/1/177/suppl/DCSupplemental

3.3 Results and discussion

3.3.1 Abundance and activity patterns of the bacterioplankton community

To quantify both the abundance and specific activity of diverse bacterioplankton in the open ocean, we sampled DNA and RNA from a well-characterized oceanographic station in the North Pacific Ocean (ALOHA) in conjunction with the Hawaii Ocean Time-series in September 2009. At this location the alpha Proteobacterium SAR11 and the cyanobacterium Prochlorococcus are generally the dominant clades throughout the year (Eiler et al. 2009; Malmstrom et al. 2010). However, the surface waters (0 - 200 m) of station ALOHA are highly stratified in key environmental variables such as
temperature, light and nitrate resulting in vertical gradients in biological properties such as chlorophyll concentration, *Prochlorococcus* and non-photosynthetic bacterioplankton abundance (Figure 6) among others. To examine the bacterioplankton community in detail, we used 16S rDNA libraries to quantify the diversity and relative abundance of specific ribotypes coupled with 16S rRNA libraries to link their phylogenetic identity to environmental activity. Large scale 16S ribosomal DNA and RNA tag libraries (mean fragment length = 390 bp) covering the variable regions V6-V8 were constructed from two depths (25 m and 100 m), which correspond to the upper photic zone and the deep chlorophyll maximum respectively (arrows, Figure 6).

Figure 6: Vertical variability of key physical, chemical and biological variables at station ALOHA. (A) Irradiance (% of surface) and nitrate (µmol kg⁻¹). Arrows indicate the depths of samples chosen for bacterioplankton diversity analysis (B)
Chlorophyll a (mg m$^{-3}$) and temperature (°C). (C) Cellular abundance measured by flow cytometry for Prochlorococcus and non-photosynthetic bacterioplankton (cells mL$^{-1}$)

A total of ~1.6 million reads (630 Mbp of sequence) were obtained among the four libraries and carefully edited for quality (see Methods) netting 861,659 sequences for subsequent analyses. In contrast with the high diversity and extreme “rare biosphere” observed by others (Sogin et al. 2006), rarefaction collector curves of each library (Figure 7) saturate (Good’s coverage ≥99%) indicating that the community diversity is well-constrained. Additionally, the number of OTUs (operational taxonomic units, here defined at 97% identity) observed in each library (219 – 575 OTUs; Table S1) is similar to values predicted for coastal seawater in a well-curated 16S rDNA clone library (Acinas et al. 2004a). Nevertheless, a substantial fraction of the OTUs from each library (16-38%) is unique to that library suggesting that further sampling may yield additional OTUs (Figure 8), even though most of the library-specific OTUs are in low abundance (<1% total). The majority of OTUs are unique to each depth (53% at 25 m, 66% at 100 m) indicating substantial microbial biogeography even in the upper 100 meters of the open ocean (Figure 8). These distinct communities form within the water column even though similar ecological and biogeochemical processes are thought to occur at both depths, albeit at much different rates.
Figure 7: Rarefaction curves for 16S rDNA and rRNA libraries from station ALOHA at 25 m and 100 m depth. OTUs were defined at 97% identity to minimize sequencing errors.
In addition to differences in the taxonomic composition between depths, the relative abundance of OTUs changes substantially both between depths and between DNA (abundance) and RNA (activity) libraries at the same depth. For example, among the 50 most abundant OTUs for the combined dataset (representing 90% of the total sequences), most OTUs are present in all four libraries (Figure 9) and they represent the broad phylogenetic groups common to many open ocean environments including *Prochlorococcus*, SAR11 and *Roseobacter* (Table S2). However, the relative abundance of these OTUs can vary substantially between depths such as for the most abundant *Alphaproteobacteria* rDNA clade (OTU 1506), which is substantially more abundant in the...
rDNA library at 25 m than at 100 m (Figure 9). OTUs can also vary in their relative rank abundance in rDNA and rRNA libraries, leading to both dramatic differences in the rRNA/rDNA ratios among OTUs within a given depth and between depths (Figure 9c). Differences in the rRNA/rDNA ratios can be further skewed by dissimilarities in the evenness of diversity among the libraries; rDNA libraries have significantly (p=0.013) higher Shannon indices (3.13 ± 0.41) compared to the rRNA libraries (1.12 ± 0.35) demonstrating a disproportionate effect of a small number of OTUs on relative activity. Due to differing numbers of copies per cell of rRNA and rDNA and the dependence of quantification on the abundance of other OTUs in the library, 1:1 ratios in rRNA and rDNA libraries would not be expected even when activity was highly coupled with abundance for a specific OTU. However, if abundance and activity are related, we would expect that the relative rank of an OTU to be the same within a given depth and in fact at both depths rRNA and rDNA rank abundance is positively related (Figure 10, Kendall’s and Spearman’s p < 0.005; 25 m Kendall’s tau= 0.1646, N=470; 100 m tau =0.0834, N=642). In spite of an overall relationship between ranks in rDNA and rRNA
Figure 9: Phylogenetic relationship of bacterial diversity with activity and abundance measurements for 16S rRNA and rDNA libraries from station ALOHA at 25 m and 100 m. (A) Maximum likelihood phylogenetic tree of the partial 16S rRNA sequences for the 50 most abundant OTUs (97% identity) in equally sub-sampled 16S rRNA and rDNA libraries (75,000 members). Bootstrap percentages greater than 80 for a given branch are indicated by a small circle. (B) Heat map showing the log abundance (log (observations+1)) for each OTU from the tree in (A) for RNA and DNA libraries. (C) Ratio of RNA to DNA (rRNA/ (rDNA+1)) showing the specific activity of each OTU for libraries from 25 and 100 m.
Figure 10: Relationship between activity (log rRNA+1) and abundance (log rDNA+1) at 25 m (open squares) and 100 m (closed circles) of each OTU. To eliminate bias, correlations are for log rRNA and log rDNA and therefore limited to OTUs where both rRNA and rDNA was present.

libraries, deviations in correspondence of relative rank between rDNA and rRNA libraries can be used to identify OTUs for which there is an uncoupling of abundance and activity. And while ~50% of the bacterioplankton assemblage is shared between
depths, the abundance (rDNA), activity (rRNA) and specific activity (rRNA/rDNA) of ribotypes varies dramatically.

The numerically dominant OTU across all libraries is most similar to *Prochlorococcus*, which is consistent with previous molecular diversity studies and flow cytometry counts (Figure 9, (Delong et al. 2006)). Despite low maximal growth rate (< 1 day⁻¹) and small cell size (Johnson and Lin 2009) that typically correspond a low ribosome requirement for cellular maintenance (Fegatella et al. 1998), *Prochlorococcus* dominates both abundance and activity libraries suggesting an important role in the ecology and biogeochemistry of the bacterioplankton community, as observed previously (Delong et al. 2006; Rusch et al. 2007). The second most abundant OTU across libraries is most similar to *Pelagibacter*, which is consistent with the SAR11 clade being one of the most abundant clades in the surface ocean (Eiler et al. 2009; Morris et al. 2002). However, this OTU, as well as other closely related alpha *Proteobacteria*, is strikingly underrepresented in the rRNA libraries (Figure 9) and in rank-order abundance is 11 positions lower in rank in the rRNA than in the rDNA libraries at both depths. This suggests that SAR11-like clades display less activity per cell than other bacterioplankton populations, with potential broader implications for their ecological and biogeochemical importance in the open ocean. This underrepresentation in libraries of activity (i.e. rRNA) may be due to a low growth rate and the extremely small cytosol volume of cells in the SAR11 clade, which is roughly an order of magnitude less than that of
Prochlorococcus cells (Sowell et al. 2008). Alternatively, these results may reflect a low point of activity in a temporally or spatially dynamic population (Campbell et al. 2011). Similar to studies in other aquatic environments (Campbell et al. 2011; Jones and Lennon 2010) we observed that OTUs often dismissed as “weeds”—low abundance species that are easily cultured and thought to be unimportant in natural environments—are disproportionately abundant in rRNA libraries. The resulting high rRNA/rDNA ratio for such “weeds” suggests that although these OTUs are not numerically abundant among the marine microbial milieu they are highly active and their populations may be regulated by tightly coupled grazing or viral pressure (Landry and Kirchman 2002; Suttle 2007), or alternately that their prevalence cycles between abundance and rarity. Interestingly, one low abundance OTU with high specific activity (rRNA/rDNA >7 ) and ranked at least 18 positions higher in the rRNA libraries than rDNA libraries at both depths is most closely related to Alteromonas, a genus shown to enhance Prochlorococcus growth under conditions of oxidative stress (Morris et al. 2011; Morris et al. 2008). Thus, this clade’s ecological niche and success, despite low abundance, could be due to a mutualistic relationship with Prochlorococcus. These results highlight that specific activity in the marine environment may be independent from abundance and that interactions between taxa in complex ocean environments may guide productivity over simple bottom-up (resource limited) or top-down (predation/viral lysis) pressures.
4.2.2 Uncoupling of specific activity and abundance in ecotypes of Prochlorococcus

Due to the abundance and biogeochemical importance of Prochlorococcus in the open ocean and the difficulty in resolving ecologically distinct Prochlorococcus clades using partial 16S rDNA gene sequences (e.g. Figure 9), we applied an alternate approach to examine this group in greater detail. Here, we used 23S rRNA targeted qPCR primers to quantitatively measure the abundance (rDNA) and activity (rRNA) of two dominant ecotypes of Prochlorococcus (high light eMIT9312/eHL-II and low light eMIT9313/eLL-IV) for two open ocean depth profiles in relation to environmental variables. In both the Sargasso Sea and the Equatorial Pacific stations, the two ecotypes exhibit maxima in cellular abundance with depth that are consistent with proposed physiological and genetic characteristics: eHL-II is adapted to high light/low nutrient regimes whereas eLL-IV is adapted to low light/higher nutrient regimes (Figure 11 panels A&D) (Moore et al. 1998; Scanlan et al. 2009). The abundances obtained with this technique are comparable to qPCR studies of ecotype abundance using well-established qPCR primers (Johnson et al. 2010b; Johnson et al. 2006; Malmstrom et al. 2010) and the most prevalent clades are captured in this analysis (Figures 6&11). While the abundance data (rDNA) aligns with the expected light-driven physiology of the clades, we see distinct results in the specific activity that differs between locations (Figure 11). Maximal specific activity largely coincides with abundance for eLL-IV in both ocean depth profiles suggesting a relatively strong coupling of population abundance and specific activity. However, the
abundance and specific activity of eHL-II are dramatically uncoupled with depth in the Sargasso Sea (but not the Equatorial Pacific) (Figure 11 panels B&E) with a maximal relative activity at ~75 m but a fairly constant cell abundance throughout the upper mixed layer (Figure 11 panels D&E). The differences in eHL-II specific activity among the locations could be explained by environmental nutrient distributions: the station in the Equatorial Pacific has nutrients at the surface where eHL-II can fully take advantage of its adaptation to high light, whereas at the station in the Sargasso Sea nutrients are below detection near the surface and eHL-II specific activity is maximal close to the nutricline even though light energy is much less favorable at that depth. Yet the abundance of eHL-II remains high near the surface in the Sargasso Sea that may be from mixing from a deeper source population, an uncoupling of growth/grazing or others processes resulting in a non-steady state condition.

As was observed in the Sargasso Sea, all of the distinguishable subpopulations of *Prochlorococcus* in the North Pacific (station ALOHA) had maximal specific activity at 100 m depth (Figure 9) even though the total *Prochlorococcus* abundance was highest near the surface (Figure 6). Thus, the observations at station ALOHA are also consistent with a deep nutricline limiting *Prochlorococcus* productivity but not cell numbers in the surface ocean (Figure 9). Regardless of the precise mechanism, measurements of specific activity (and not abundance) are most consistent with the observed nutrient and light distributions. Thus specific activity measurements identify locations or time points
when a resource limitation-only model (bottom up control) for *Prochlorococcus*
populations (and likely other bacterioplankton) does not apply and provides a potential
methodology for resolving unexplained biogeography or other ecological patterns in
genetically distinct groups of bacterioplankton (Fuhrman 2009).

Figure 11: *Prochlorococcus* abundance and activity with associated
physicochemical variability versus depth for two representative locations in the
Equatorial Pacific (A-C) and Sargasso Sea (D-F). Abundance profiles (A,D) of the two
dominant ecotypes high light eMIT9312/eHL-II and low light eMIT9313/eLL-IV
measured using qPCR of 23S rRNA genes (cells mL⁻¹). Normalized specific activity
(B,E) of eHL-II and eLL-IV ecotypes measured using qPCR of 23S rRNA and rDNA.
Profiles of physicochemical variability (C,F) including irradiance (% of surface), temperature (°C) and nitrate (µmol kg⁻¹).

4.2.3 Implications of uncoupled specific activity and abundance

Marine microbiology has progressed dramatically from assessing the abundance and diversity of the bacterioplankton (16S rDNA surveys), to functional assignments using genomics and metagenomics. Now, there are methods that can relate the presence of an organism or gene in relation to its expression or activity in that environment (rRNA RT-qPCR, metatranscriptomics) (Gifford et al. 2011). Our results suggest that low abundance microbes may be disproportionately active in certain environments (Alteromonas) and that some of the most abundant (e.g. SAR11) may have low metabolic activity. Comparisons of rRNA and rDNA levels in situ have the potential to estimate environmental growth rates and activity without the biases associated with incubations or culturing. Although absolute quantification of a specific target is not possible in gene libraries, RT-qPCR of rRNA (cDNA) and qPCR of rDNA provide estimates of specific activity for clades of organisms and are complementary to library-based approaches that allow inter-comparison among many groups. We observed uncoupling of abundance and specific activity of Prochlorococcus in the Sargasso Sea depth profile, which highlights deficiencies in our understanding of marine microbial ecology and population structure. If the specific activity is higher for eHL-II deeper in the water column, was this population in the process of becoming more abundant or do other
factors such as predation limit their abundance at depth? Techniques that allow us to investigate the activity and abundance for specific ecotypes in situ will allow us to examine the environmental factors that structure bacterioplankton populations and link the relative abundance and specific activity of marine bacterioplankton populations. Future studies should include a broad suite of methodologies that assess different metrics of microbial activity (e.g. ATP, carbon uptake, growth rate) to compare how rRNA-based activity is related to these other measurements.

3.4 Acknowledgements

This work was funded in part by funds from the US National Science Foundation (#OCE10-31064). Sequencing was conducted in part by the U.S. Department of Energy Joint Genome Institute and is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors would like to thank Brett Updyke for assistance with sampling and Brian Binder for valuable discussion. We also acknowledge helpful comments from the anonymous reviewer.
4. The Temperature Dependence of the Activity of Genetically Diverse Prochlorococcus

4.1 Introduction

A major effect of global climate change is the increase of mean sea surface temperature (Meehl et al. 2007). As a key parameter of thermodynamics, temperature regulates biochemical reactions and enzyme kinetics, which are fundamental to all biological processes. From unicellular microbes to multicellular animals and plants, temperature has been shown to be one of the primary determinants of general metabolic rates (Gillooly et al. 2001) which contribute to all other biological activities. Even though temperature increases may result in increased general metabolic rates, the kinetics of various biochemical reactions may respond to temperature differently and thus changes in temperature may alter the balance between different metabolic activities such as production and respiration. Furthermore, over evolutionary time scales, different organisms have adapted to particular temperature ranges and unexpected environmental shifts in temperature may decrease the abundance or even lead to localized extinction of less tolerant species thus alter the community composition. For
marine microbes in particular, whose activities directly drive essential ecological and biogeochemical functions in the ocean, there is an urgent need to reveal how changes in temperature (along with other environmental factors) can affect their metabolic activities and its impact on microbial community structure in order to predict the consequences of climate change.

The marine cyanobacterium *Prochlorococcus* is an excellent model organism for studying the effects of climate change because they are relatively genomically streamlined (~ 2000 genes at minimum), yet, have widespread dominance (numerically) in global surface oceans. They inhabit tropical to subtropical open oceans, ranging from 40°S to 48°N (Johnson et al. 2006; Partensky et al. 1999), and their photosynthetic activity has been estimated to account for 25 to 45% of primary production in oligotrophic oceans (Goericke and Welschmeyer 1993; Jardillier et al. 2010). Based on the differences in genetic and physiological characteristics (Moore and Chisholm 1999; Moore et al. 2005; Palenik and Haselkorn 1992; Rocap et al. 2002) and biogeography distributions (Johnson et al. 2006; West et al. 2010; West and Scanlan 1999), *Prochlorococcus* has been divided to different ecotypes which are hypothesized to occupy different environmental niches. Two major groups are described as high-light (HL) and low-light (LL) adapted types according to their different optimal growth irradiance (Moore and Chisholm 1999; Moore et al. 1998). In their natural environmental, it has been observed that the HL clade dominates the surface of the water column while the LL clade is most abundant at
depth (Johnson et al. 2006; Ralf Goericke 1993; Zwirglmaier et al. 2008). The HL group further splits into at least two clades, eHL-I and eHL-II, with eHL-II dominates the surface ocean in the 30N – 30S and eHL-I thrives at higher latitude (Johnson et al. 2006; Zinser et al. 2007; Zwirglmaier et al. 2008). Laboratory growth experiments using representative strains from these two ecotypes reported that eHL-I strains outgrew eHL-II strains at lower temperature and vice versa (Johnson et al. 2006). The observation of temperature being correlated with shifts in field distributions as well as temperature affecting growth rates in lab lead to the hypothesis that temperature is the primary environmental variable responsible for observed differences in relative abundances of eHL-I and eHL-II. To test this hypothesis as well as to further understand the effects of temperature on the ecology and function of these two clades, the following questions need to be addressed: besides growth rate how do other essential cellular activities such as photosynthesis, respiration, and ATP production vary as eHL-I and eHL-II exposed to temperature gradients? Do complex environmental factors (such as nutrients, mixing, and oxygen) affect their response to temperature? How does the relationship between activity and abundance affected by temperature in the open ocean?

In order to further investigate the response to temperature of Prochlorococcus clades eHL-II and eHL-I, we conducted both laboratory and field based experiments to investigate how activity changes in representative strains as a function of temperature. In the lab under relatively short-term acclimation, strains MIT9312 and MED4 were
cultured at different temperature representing the latitudinal range of *Prochlorococcus*.

Different metrics of key metabolic activities, including specific growth rate, rRNA content, primary production (photosynthesis), electron transport system (respiration), and ATP content were estimated as a function of temperature. To observe how temperature may affect activity in natural environments, ecotype specific biomass and activity (rRNA based) of coexisting eHL-II and eHL-I were characterized along a strong sea surface temperature gradient from the Central Pacific Gyre (Hawaii) to California Current system. Based on previous culture studies and field biogeography patterns (Johnson et al. 2006; Zinser et al. 2007), we hypothesize that eHL-II is more active than eHL-I in warmer water and vice versa.

**4.2 Experimental Procedures**

**4.2.1 Culture conditions**

Axenic strains MED4 (eHL-I) and MIT9312 (eHL-II) were grown in 75% filtered Sargasso Sea water (0.2 µm) amended with Pro99 nutrients (Moore et al. 2002). For measuring activity as a function of temperature, 20 ml batch cultures were grown in triplicate in an aluminum temperature-gradient bar on a 14:10-h light: dark cycle with constant day light at 66 ± 10 µE m⁻² s⁻¹. MED4 and MIT9312 were cultured in two separated temperature bars, each with light incident from below using cool white
fluorescent lamps adjusted with blue stage screening. *In vivo* chlorophyll fluorescence was monitored at the same time each day using a Turner 10-AU fluorometer (Sunnyvale, CA) and cultures were transferred to fresh media before entering stationary phase. Growth rates were calculated by linear regression of natural log transformed fluorescence versus time. After acclimating to each temperature for at least 7 days, cells in exponential phase were harvested in the incubator afternoon (5 h before lights off).

4.2.2 Flow cytometry

For flow cytometry analysis, 0.5 ml of culture was preserved with 0.125% glutaraldehyde and frozen at -80°C until later analysis. *Prochlorococcus* cells were enumerated using a FACSCalibur flow cytometer (Becton Dickinson) as previously described (Johnson et al. 2010b). Population properties (scatter and fluorescence) in geometric mean were normalized to 1.0 µm yellow green polystyrene beads (Polysciences, Warrington, PA).

4.2.3 Primary production (photosynthesis activity)

Primary production was estimated from $^{14}$C labeled bicarbonate incorporation during 2 h light incubation as previously described (Barber et al. 1996) with several modifications for measuring lab cultures. Culture was dispensed into four borosilicate scintillation vials (0.5 ml each) and inoculated with 10 µCi of NaH$^{14}$CO$_3$ solution. Those vials were then incubated for 2 h in an environmental chamber with light and temperature set the same as those the culture acclimated to in the temperature bar, with
two of the four vials wrapped in aluminum foil to provide no-light controls.

Meanwhile, to assess the total radioactivity added to each sample, two blank vials were inoculated with 10 µCi of NaH\(^{14}\)CO\(_3\) solution, one drop of beta-phenylethylamine and 3 ml of Ecolume scintillation cocktail (ICN Biomedicals, Costa Mesa, CA). After incubation, samples were preserved and acidified by adding 50 µl of 37% formaldehyde and 50 µl of 1N HCl. The vials were then left uncapped to degas overnight to remove un-incorporated labeled substrate. Ecolume scintillation cocktail (3 ml) was added the following day, and the radioactivity was quantified. Primary production was estimated from the activity of the total and each sample following standard procedures (Barber et al. 1996).

4.2.4 Electron transport system (ETS) activity

ETS activity, which indicates total respiration potential, was assessed by quantifying the enzymatic oxidation of NADH and NADPH in the presence of Iodonitrotetrazolium (INT) as the electron acceptor and an indication dye (Packard et al. 1971; Span 1986). The original method (Span 1986) was optimized in this study to yield more robust reading for *Prochlorococcus* cultures.

Extraction solution A consists of MgSO\(_4\) (150 mM), Polyvinyl pyrolidone (2 g l\(^{-1}\)) in Tris-EDTA buffer (50 mM Tris, 5 mM EDTA, pH = 7.4). Extraction solution B consists of Triton X-100 (38 ml l\(^{-1}\)) in Tris-EDTA buffer. Substrate solution consists of sodium succinate (133 mM), NADH (3.4 mM), NADPH (1 mM) in Tris-EDTA buffer.
Termination buffer consists of 1:1 (by volume) 37% formaldehyde and 1M phosphoric acid. All solutions above were kept at 4°C. Incubation solution consists of Iodonitrotetrazolium chloride (INT) (2 g l⁻¹) in H₂O and was stored at -20°C.

First, 9 mL of fresh culture (~10⁷ cells ml⁻¹) was filtered onto a 25 mm 0.2 µm polycarbonate membrane, and placed in a vial with 100 µl of extraction buffer A. The sample was then shaken without beads in a bead-beater (BioSpec, Bartlesville, OK) at 4,800 rpm for 3 min with cooling on ice for every 1 min. After adding 100 µl of extraction buffer B and vortexing, the sample was centrifuged for 2 min at 20,000 g. The supernatant (enzyme extract) was split into two tubes (50 µl each) with one heated at 95°C for 3 min to create an enzyme blank. Both tubes were then incubated with 50 µl of INT solution and 150 µl of substrate solution at room temperature for 30 min. Termination solution (50 µl) was added to stop the reaction, and the absorbance at 490 nm was determined using a Synergy4 microplate reader (Bio-Tek Instruments, Inc., Winooski, VT). Relative ETS activity is calculated as the absorbance at 490 nm minus the boiled blank control.

4.2.5 ATP

ATP, the molecular unit of currency of intracellular energy transfer (Knowles 1980), was quantified based on the firefly bioluminescence reaction using CellTiter-Glo kit (Promega, Madison, WI). The experimental procedures were modified from the manufacturer’s instructions to optimize the reaction for Prochlorococcus culture. Briefly,
duplicate wells in a 96-well microplate were loaded with 50 µl of CellTiter-Glo reagent (containing luciferase and buffer) and 100 µl of fresh culture (~10^7 cells ml^-1) in parallel with serially diluted ATP standards ranging from 0.1-200 nM. Two wells were loaded with 100 µl of culture media as blank control. The mixtures were then incubated at 38°C for 5 min and the luminescence at 560 nm was measured in a Synergy4 microplate reader (Bio-Tek). Sample ATP concentration was determined by comparing the luminescence signal to standards.

4.2.6 RNA extraction and RT-qPCR

For lab sampling, 0.5 ml of fresh culture was filtered onto a 25 mm 0.2 µm polycarbonate filter and stored in liquid nitrogen immediately. Right before RNA extraction, 650 µl of Qiagen RLT solution (1% 2-mercaptoethanol pre-added) was added to each sample. For environmental sampling, 100 ml of seawater was filtered and the membrane was directly preserved in 600 µl of RLT solution with 1% 2-mercaptoethanol and stored in liquid nitrogen.

RNA extraction and reverse transcription were performed following Procedure C as described (Lin et al. 2013). After cDNA synthesis, strain MIT9312 or ecotype eHL-II 23S rRNA was quantified by qPCR as previously described (Johnson et al. 2006). Strain MED4 or ecotype eHL-I 23S rRNA was quantified using modified qPCR primers 23S-1194F (5’-GAAGCTGCGGATAAATTTATTT-3’) and 23S-1443R (TCCCCCTTCCCCAAAACATAA-3’) at a final concentration of 500 nM. The qPCR
program was 95°C for 15 min, 45 cycles at 95°C for 15 s, 64.3°C for 30 s, and 72°C for 15 s, followed by a final extension of 5 min at 72°C. The specificity of the primers was tested using representative *Prochlorococcus* and *Synechococcus* isolates as well as clone libraries constructed from environmental samples.

**4.2.7 Field sample collection and environmental characterization**

Samples from the Central North Pacific Gyre to the California Current were collected during 1 March - 10 March 2012 aboard the R/V *Thomas Thompson*. Surface seawater from the vessel seawater supply line was sampled twice daily (4:00-5:00, 16:00-17:00, UTC time) to collect DNA, RNA and flow cytometry samples. For DNA and RNA, 100 ml of seawater was filtered and preserved as previously described (Lin et al. 2013). Temperature data was collected from the ship’s underway seawater flow-through analysis system. In addition, hydrographic data and water samples were collected daily (around 15:00 UTC time) using a CTD rosette system. Chl *a* was extracted on board and quantified using a Turner 10-AU fluorometer as described in (Johnson et al. 1999). Samples for macro-nutrients (including NO₃ + NO₂, PO₄, SiO₂) were stored at -80°C and later quantified on land using an Astoria-Pacific A2 autoanalyzer according to the manufacturer’s protocol. The concentration of NH₄ was determined on board with a Turner TD700 fluorometer based on a fluorescence method (Holmes et al. 1999).
4.3 Results and discussion

4.3.1 Culture growth

Strain MIT9312, a member of the eHL-II clade, and strain MED4, a member of the eHL-I clade, both had maximal growth rates between 24 to 25 °C, and the crossover point of their growth curves was around 18.6°C – above this temperature MIT9312 grew faster than MED4 and below this point MED4 outgrew MIT9312 (Fig. 12A). Whereas the shape and general trend of the temperature – growth relationship are consistent with an earlier study (Johnson et al. 2006) and phytoplankton responses to temperature (Thomas et al. 2012), we observed lower maximum growth rate (0.47 d⁻¹ for MIT9312 and 0.31 d⁻¹ for MED4) and narrower temperature tolerances (18 - 30 °C for MIT9312 and 16 - 27 °C for MED4) than previously described. This may be the result of using axenic cultures in this present study, which is necessary for many of the activity estimates. Previously, it has been reported that heterotroph bacteria could facilitate Prochlorococcus growth (Morris et al. 2008). Alternatively, the low salinity media (75%-seawater based) may also contribute to the growth difference.

Side scatter (SSC), which reflects cell biovolume and cellular composition (Ackleson and Spinrad 1988), is positively correlated with both temperature and growth
rate (Fig. 12B). This trend is consistent with the classic cell size – growth relationship described in model heterotrophic bacteria, but is in contrast to the results of *Prochlorococcus* light-limit growth experiments (Lin et al. 2013; Worden and Binder 2003b).

These distinct size-growth responses suggest that light and temperature may influence *Prochlorococcus* growth rate through different mechanisms: within certain range temperature regulates cell growth through thermodynamics, and increased metabolism at higher temperature could lead to faster biomass accumulation (replicating DNA, RNA and proteins) thus bigger cells; light constrains cell growth through photosynthetic energy availability, and at lower light level cell size increases due to packing of...
additional photosynthetic apparatus to more efficiently harvest light (and relatively less investment in growth). While the temperature mechanism is probably conserved between heterotroph and autotroph, the light mechanism is autotroph specific. Furthermore, except for the 18.2°C point, MIT9312 in general is larger than MED4 perhaps due to the higher growth rates of MIT9312 or strain specific physiology (i.e. MIT9312 was isolated from a deeper depth (Moore et al. 1998)).

4.3.2 Activity Metrics as a Function of Temperature

Primary production (photosynthesis) (Fig. 13A) ranged from 4 - 10 fg C hr\(^{-1}\) cell\(^{-1}\) for MED4, and 4 - 18 fg C hr\(^{-1}\) cell\(^{-1}\) for MIT9312. Both strains exhibit a single peak, with the optimal primary production temperature at about 28°C for MIT9312 and about 24°C for MED4. At temperatures below the maximal production value, the two strains respond identically and the upper bound forms a continuously increasing trend from 16 - 28°C, which resembles the widely cited Eppley curve (Eppley 1972) (i.e. temperature sets an upper limit on exponentially increasing phytoplankton photosynthesis and growth rate). This phenomenon suggests while biogeochemical contribution (in this case primary production) of Prochlorococcus is strongly influenced by temperature, it may be independent of the community composition of Prochlorococcus, i.e. MIT9312 and MED4 could be functionally redundant over certain temperature ranges.
Figure 13: Variation of cellular activities (A) primary production per cell, (B) rRNA per cell, (C) ATP (pmol) per cell and (D) ETS activity per cell across a gradient of temperature for MIT9312 (red triangles) and MED4 (blue circles).
While primary production varies as a function of temperature, both cellular rRNA and ATP content were relatively constant within the temperature tolerance range of each strain (Figs 13B and C). MIT9312 contained significantly higher levels of RNA and ATP per cell than MED4 regardless of temperature (T-test, P<0.01). ETS activity, the enzymatic oxidation of NAD(P)H, is a key process involved in both respiration and cyclic electron transport of photosystem I in Prochlorococcus (Battchikova and Aro 2007; Zinser et al. 2009). In general, ETS pattern is quite variable compared to other activity metrics (Fig. 13D). MED4 had an ETS optimal at 21.9°C, and except for this point MIT9312 showed overall higher activity than MED4 (T-test, P<0.01). The error bars for ETS are bigger than other activity estimates, likely due to the low biomass of Prochlorococcus (9 ml of culture, \( \sim 10^8 \) cells) used in this study. Increasing filtration volume could help to improve the accuracy of this assessment for Prochlorococcus cells.

Although the magnitude of change and the optimal temperature differs among the metrics of activity, the shape of each activity – temperature relationship is similar between MIT9312 and MED4. This suggests that those two genomically distinct strains may share the same basic mechanisms in response to temperature, albeit with different thresholds or maxima. Comparing the different metrics of activity, a nonparametric analysis revealed three significant activity correlations: 1) specific growth rate and photosynthesis per cell for strain MIT9312 (Spearman correlation coefficient = 0.62, P<0.01) and strain MED4 (Spearman correlation coefficient = 0.86, P<0.01) as well as the
two strains combined (Spearman correlation coefficient = 0.78, P<0.01); 2) cellular rRNA and ATP content for the combined dataset of the two strains (Spearman correlation coefficient = 0.79, P<0.01) but not significant for either strain individually (P>0.1); and 3) cellular rRNA and specific growth rate for strain MED4 (Spearman correlation coefficient = 0.78, P<0.01) and the two strains combined (Spearman correlation coefficient = 0.52, P<0.01) but not significant for strain MIT9312 (P>0.1). These correlations suggest that there are interactions between different metrics of activity. Among the measured activities, growth rate and primary production are directly related to ecological and biogeochemical functions in the field, and according to our results both activities are highly sensitive to changes in temperature.

4.3.3 Relationship of rRNA content and growth rate

The rRNA-growth relationship is of particular interest to researchers because rRNA could be potentially developed as an indicator for in situ bacterial activity. Our previous work (Chapter 2) showed that under light-limited conditions there is a strong positive linear correlation between rRNA cell\(^{-1}\) SSC\(^{-1}\) and specific growth rate for both strains MIT9312 and MIT9215, which are members of the eHL-II clade. However, under temperature regulation, the rRNA content – growth relationship appears to be strain (or clade) specific and is not described by a single relationship (Fig. 14). In particular, the pattern for MIT9312 does not follow the overall trend of increasing rRNA cell\(^{-1}\) as a function of growth rate and instead has a relatively constant cellular rRNA throughout
the range of temperature-limited growth rates (Fig. 14A). Furthermore, after normalized by size (SSC), there are significant but opposite linear relationships between the two strains for rRNA cell$^{-1}$ SSC$^{-1}$ and growth rate (Fig. 14B), i.e. positive for MED4 ($R^2 = 0.81$) and negative for MIT9312 ($R^2 = 0.80$). This suggests that temperature affects the RNA-growth relationship differently between strains (MED4 and MIT9312) and possibly clades.

For MIT9312, why is there a strong positive rRNA-growth relationship under variable light but no correlation (rRNA cell$^{-1}$ vs. growth rate) or negative correlation (rRNA cell$^{-1}$ SSC$^{-1}$ vs. growth rate) under variable temperature? One possible clue is that the two MIT9312 regression lines (Fig. 14B) crossed at about 22°C of the temperature-regulated line (Fig. 14B) which is the constant temperature used in the light regulated growth rate experiment (Chapter 2). When temperature is <22°C, temperature-regulated MIT9312 cells held more rRNA per biomass than the light-limited cells at the same growth rate even though the temperature-regulated cells were smaller under saturated irradiance. When temperature is >22°C, temperature-regulated cells accommodated less rRNA per biomass compared with light experiment. One possible explanation is that ribosome efficiency increases at higher temperature. According to (Bremer and Dennis 1996b), bacterial growth rate can be described by a theoretical relationship:
\[ \mu = \left( \frac{60}{\ln 2} \right) \times \left( \frac{Nr}{P} \right) \times \beta r \times cp \]

(1)

Where \(Nr\) = ribosomes per cell; \(P\) = protein per cell; \(\beta r\) = the fraction of active ribosomes; \(cp\) = peptide chain elongation rate (or protein production rate per ribosome).

In \textit{E. coli} temperature has been shown to strongly influence the peptide chain elongation rate, e.g. 3.2-fold increase of \(cp\) from 20 to 40°C (Ryals et al. 1982). This temperature - peptide chain elongation rate relationship likely resembles the classic temperature driven enzyme kinetics curve, i.e. positive increasing reaction rate (but not necessarily linear) with temperature until a maximum is reached, beyond which the rate declines.

For a given rRNA concentration (rRNA per biomass), higher temperature could enhance protein synthesis rate and therefore leads to the higher observed growth rates. This temperature – ribosome efficiency effect may explain the negative slope of rRNA cell\(^{-1}\) SSC\(^{-1}\) growth regression line for MIT9312 and lower slope (compared with the light experiment) for MED4, i.e. with higher ribosome efficiency at higher temperature fewer ribosomes are needed at the same growth rate. Alternatively, temperature does not appear to regulate cellular rRNA concentration in \textit{E. coli} (Farewell and Neidhardt 1998; Ryals et al. 1982) or \textit{S. typhimurium} (Schaechter et al. 1958). Those findings indicate that rRNA concentration and temperature both regulate growth but should be treated as two independent variables. According to equation (1) it is possible to parameterize a growth rate function with measured rRNA content and temperature; however, this calculation
requires that the relationship between \( cp \) and temperature has to be quantitatively resolved.

Figure 14: (A) rRNA cell\(^{-1}\), and (B) rRNA cell\(^{-1}\) normalized by SSC as a function of growth rate for MIT9312 (red triangles), MED4 (blue circles) and previous light experiment results (black crosses).

To predict growth, an alternative method is empirical fitting with rRNA and temperature simply treated as independent linear variables (Buckley 1984). We combined the light and temperature experiments (n = 22), and performed a stepwise linear regression using rRNA cell\(^{-1}\) SSC\(^{-1}\) (R) and temperature (T) to explain growth rate. This analysis gave an empirical function:

\[
\mu = 2.746 \times 10^{-3} \times R + 0.02 \times T - 0.35 \quad (16 < T < 30^\circ C)
\]

Including results from three strains (MIT9312, MIT9215 and MED4) and two different experiments, this preliminary model could explain a substantial fraction of the
variability in growth rate (R^2 = 0.78), while as a single factor rRNA cell^-1 SSC^-1 explains only 65% and temperature alone only explains 11% of the variability. The standard error of the estimate (the square root of the residual mean square) is 8.4%. More data points with nested temperature and light manipulation would help to improve the predictive power of this model. The robustness has to be tested by independent growth experiments, and more Prochlorococcus strains should be investigated especially from more evolutionarily distant low-light ecotypes.

4.3.4 Field patterns across a natural sea surface temperature gradient

To explore the ecotype specific activity patterns for eHL-I and eHL-II as well as to investigate the temperature effect in a natural setting, we sampled the surface water from Hawaii to San Diego in March 2012 (Fig. 15A). As expected from latitudinal effect and oceanic currents, this transect crossed a strong gradient in sea surface temperature (13.8 – 22.1°C) that based on laboratory experiments are expected to influence the activity, including growth rate, of Prochlorococcus. Compared with depth profiles, the irradiance variability is relatively small along this horizontal transect since sampling focused on near-surface depths. This approach provides an opportunity to examine the temperature effect independent of light change, both of which have been shown to strongly influence Prochlorococcus growth and production (Moore and Chisholm 1999; Vaulot et al. 1995)(Chapter 2). However, the two ends of this transect represent distinct oceanographic regions (i.e. the Central North Pacific Gyre and California Current
Figure 15: (A) Sample locations and sea surface temperature (SST in °C) in the North Pacific Ocean. Surface transect profiles of (B) biomass distribution, (C) rRNA per cell for Prochlorococcus ecotypes eHL-II or eMIT9312 (red triangles) and eHL-I or eMED4 (blue circles), and (D) environmental variables including SST, PO4 (red dashed line), mixed layer depth (black squares) and percentage of small size phytoplankton chlorophyll (0.2 – 2 µm). Grey arrows highlight two key transition points.
system), which are significantly different in many environmental variables including nutrients (N, P, Si and Fe), pH, oxygen and phytoplankton community composition.

As assessed by qPCR, two high-light ecotypes differentially dominate the surface water in different regions. Whereas eHL-II thrived in warmer waters, eHL-I outnumbered eHL-II in colder areas, with a crossover point at 131.7°W (T = 18.9°C) (Fig. 15B). The summed surface abundance of those two ecotypes was in general ~10^5 cells ml^-1 but dropped sharply at the three coldest stations near the coast. These biomass distributions along the temperature gradients and oceanic-coastal transition are consistent with previous observations in the Atlantic Ocean (Johnson et al. 2006; Zinser et al. 2007; Zwirglmaier et al. 2007). As shown in Fig. 15C, cell size (SSC) was relatively constant (CV = 4.8%) over sampled regions, except for the last coastal station where it increased substantially by 55%. This increase is likely due to enriched nutrient concentrations (Fig. 15D) as previous culture studies (Bertilsson et al. 2003) reported significantly higher carbon content in nutrient-replete MED4 cells compared with nutrient-limited cells.

To reveal ecotype specific activity, using RT-qPCR we estimated the cellular rRNA content of the two dominant clades, eHL-II and eHL-I, with ecotype specific primers targeting the 23S gene. From west to east, except for the last three stations, in general eHL-II rRNA cell^-1 was higher than eHL-I in warmer water while eHL-I contained more rRNA cell^-1 in colder water (Fig. 15C), which is consistent with the trend
in their relative biomass. The rRNA cell\(^{-1}\) for eHL-II and eHL-I crossed at 132.7 °W with temperature of 19.4°C which was slightly higher than the biomass crossover temperature. The estimated rRNA content for eHL-I and eHL-II ranged from 20 – 240 copies cell\(^{-1}\). The variability of rRNA patterns may partially be due to the dynamic nature of activity as compared to biomass as well as potential diel variations (nucleic acid samples were taken in the morning and evening) (Chapter 2). At the three stations east of 125.3°W, particularly high eHL-I rRNA cell\(^{-1}\) were estimated, ranging from 420 – 830 copies cell\(^{-1}\) (i.e. about 4 – 8 fold increase compared with the 125.3°W point). Size alone (at the maximum increased by 50%) could not account for this dramatic rRNA increase and cDNA clone library indicated no sign of nonspecific RT-qPCR amplification (data not shown). The temperature at these stations (13.8 – 15.4°C) was on their threshold of survival temperature for our axenic laboratory cultured MED4. One possible explanation is that, in response to low ribosome efficiency at cold temperature, wild type eHL-I (potentially different subclade from open ocean strains) has evolved to accumulate high level of cellular rRNA to maintain cellular activity and the nutrient enriched coastal environment could further facilitate this accumulation (i.e. rRNA synthesis requires high P, N). Isolates from low temperature and coastal adapted eHL-I strains would help to address this hypothesis.

Recognizing potential limitations, we applied equation 2 to estimate the growth rate using rRNA content, cell size (SSC) and \textit{in situ} temperature. The three stations at
the east end were excluded from this analysis because 1) the temperatures were below the laboratory survival range based on which the empirical equation that was established and 2) the rRNA contents were dramatically higher than other stations perhaps reflective of different physiology in adaption to the coastal environment. The cell size of eHL-II or eHL-I was assumed the same as the mean size of the total *Prochlorococcus* community as no significant subpopulation differentiation indicated by the SSC histograms. It should be noted that because of sampling frequency our SSC data could not eliminate diel variation. The estimated potential growth rates, as illustrated by Fig. 16, ranged from 0.2 – 1.0 d\(^{-1}\), which is in general higher than lab isolates but comparable to previously reported field growth rates of *Prochlorococcus* at the community level (Goericke and Welschmeyer 1993; Landry et al. 2008; Liu et al. 1999; Vaulot et al. 1995). The overall trends highly resemble the rRNA cell\(^{-1}\) patterns (Fig. 15C) and are inconsistent with observed biomass transition between eHL-II and eHL-I (Fig. 15A). Moreover, compared with the laboratory-based MED4 growth curve with monotonic decrease in growth rates as temperature dropped from 24°C to 16°C (Fig. 12A), the estimated eHL-I potential field growth rates shows a significant increase around 132°W (20 - 18°C). This mid-temperature peak of eHL-I rRNA content/ potential growth rate seems to be a response to other environmental variables, such as decreasing mixed layer depth (thus higher time averaged irradiance) (Fig. 16) or elevated nutrient concentrations (Fig. 15D).
Combining of biomass, rRNA cell$^{-1}$ predicted growth rate, and environmental measurements together, three distinct zones emerge with two key transition points at ~132°W and 124°W, respectively (Fig. 15). The west zone is dominated by eHL-II both in biomass and rRNA content, and the water is characterized by high temperature (>19°C), low nutrients and deep winter mixing. The central zone is dominated by eHL-I both in biomass and rRNA content, and the temperature is between 19 - 16°C with elevated nutrients and shallow mixing. In the east zone where temperature is low (<15°C), nutrients are high and phytoplankton community dominated by larger cells, the biomass of both eHL-II and eHL-I drops rapidly even though eHL-I rRNA content is
high. At each transition point, besides sea surface temperature, environmental variables such as $\text{PO}_4$ (representing a suite of co-varying macro-nutrients) and mixed layer depth exhibit significant transition point as well (Fig. 15D). Other biotic and abiotic variables such as iron, oxygen, pH, and heterotrophic bacteria may also contribute to the transitions, and this is likely related to physical oceanic fronts separating different water mass.

### 4.4 Conclusions

Using *Prochlorococcus* clades eHL-I and eHL-II as a model system in the context of temperature as a major driver of ecological differentiation, we investigated how activity changes along a well-controlled laboratory temperature gradient as well as natural temperature gradients in the field. In the lab, five different activity metrics as a function of temperature were estimated, including specific growth rate, primary production, ATP, rRNA contents and ETS activity. Whereas strains MED4 and MIT9312 demonstrate different activity level and optimal temperatures, the general trends are similar indicating conserved temperature regulation mechanisms between these two strains. Growth rate and primary production, which are two activities directly related to ecological and biogeochemical functions, are most sensitive to temperature change among measured activities. Furthermore, based on the results of this temperature
experiment and previous light experiment, an empirical function was developed to estimate potential growth rate using rRNA content and temperature. In the field, we observed transitions between eHL-I and eHL-II along a strong temperature gradient. Based on the patterns of biomass, rRNA contents and estimated potential growth rates, three distinct zones and two transition points were identified. Temperature alone could not explain all the field variability; and the transition areas were characterized by steep environmental gradients in multiple key environmental variables such as nutrients and vertical mixing. Future studies should generate more in situ activity data based on multiple metrics, and further investigate the environmental effects particularly at critical transition zones. Under temperature regulation, how do associated environmental variables influence activities and community compositions? Or in a broader context, how does the marine ecosystem as a complex system respond to climate change?
Conclusions

Using Prochlorococcus as a model system, this study presents a body of work to develop a pioneering approach to quantify the in situ activity of ecologically distinct, genetically defined clades of marine bacteria. Based on the methods we developed in lab, ecologically meaningful activity patterns have been revealed in the field. This unique information, as a complement to existing diversity and abundance assessments, will help to unlock the ‘microbial black box’.

In the laboratory, we have shown that specific growth rates of representative Prochlorococcus strains could be quantitatively predicted from cellular rRNA content, cell size and temperature. Two key environmental factors, light (Chapter 2) and temperature (Chapter 4), were used to regulate Prochlorococcus growth in lab. For ecotype eHL-II, strain specific rRNA-growth relationship has been observed (MIT9312 vs. MIT9215), but when size is accounted for the differences between the two strains are eliminated (Chapter 2). This is encouraging because it suggests a potential universal function of rRNA vs. growth may exist. Size normalized rRNA contents (rRNA cell$^{-1}$ SSC$^{-1}$), indicating cellular protein synthesis potential, explains most of the growth variance (92%) under light-limit condition, and significant amount of variance (65%) under both light and temperature manipulations. One hypothesis that emerged from the two datasets: limited resources such as light (or potentially nutrients) may directly
constrain rRNA concentration and thus impact growth whereas temperature may regulate growth through changes in the ribosome efficiency, which in turn is independent of rRNA content. Taking rRNA, cell size and temperature together, an empirical model could explain 78% growth variance including both light and temperature experiments for three strains from two different ecotypes (eHL-I and eHL-II) (Chapter 4). This work sets up an experimental framework to examine the quantitative relationships of rRNA – growth in a laboratory system, which could be extended to other isolated microbes representing key environmental or human health functional groups. For the Prochlorococcus system, the robustness of the proposed model should be further tested using more lab isolates in particular investigating strains that are more evolutionary distant such as members of the low light clade. To improve the prediction power, the effects of other environmental variables should also be examined. Resource limited factors such as macronutrients may resemble the light effects (i.e. rRNA concentration dependent growth regulation); and it is unclear how other stress factors such as oxygen and salinity may impact rRNA concentration and growth. Alternatively, the model can be directly tested in the field (under natural combination of complex environmental factors) using activities assessed by other independent methods, such as cell division frequency (diel FCM) and primary production. A preliminary comparison of eHL-II rRNA and community primary production was performed at two field stations and showed excellent coherence (Chapter 2).
While there may be further refinements for this molecular approach to estimate in situ activity, our field studies represent the first report of *Prochlorococcus* ecotype activities within natural mixed populations. In vertical profiles, estimated growth rates for eHL-II, a dominant high-light ecotype, are strongly influenced by light (Chapter 2) which is consistent with their physiology known from lab isolates. In addition to light, we show that nutrients and mixing also contribute to the variance of eHL-II activity within the water column. Compared with relatively constant biomass distribution within the mixed layer, activity patterns are more dynamic and demonstrate significant subsurface peaks. The potential mechanisms causing the decoupling between biomass and activity patterns, such as selective grazing, virus lysis and vertical mixing, need to be further investigated in future studies. In addition to strong vertical gradients, there are major horizontal gradients both zonally and meridional albeit typically at much larger spatially scales. Across a strong natural temperature gradient from Honolulu to San Diego, we estimated clades specific activities for two high-light adapted ecotypes eHL-I and eHL-II (Chapter 4), which has been hypothesized to partition surface waters because of different temperature niches. Our results show that their activities indeed differentiate along the temperature gradient, with higher eHL-II activity in warm water and higher eHL-I activity in cold water; and the activity transition is more or less consistent with their biomass transition. However, when looking in more details, significant changes of environmental variables such as nutrients and mixed layer depths
were identified right at the transition point which may also contribute to the niche
differentiation between eHL-I and eHL-II in the field. Is it temperature alone or a
combination of environmental factors defining the niches for these two ecotypes? If it is
a combination, how should we parameterize these different environmental variables?
More field data are needed to elucidate the correlation between the observed variability
in in situ activity and environmental variables.

Broader application of this rRNA-based activity in the field intensively will
generate a wealth of data, and based on those data, 1) key environmental drivers should
be identified or tested, 2) the growth of major environmental functional groups should
be parameterized, and 3) these in situ taxon-specific parameterization will be used to
improve current marine ecosystem models (Follows et al. 2007).

Using ribosomal tag pyrosequencing of DNA and RNA, we have extended our
observation of taxon-specific activity to the Eubacterial community. Our analysis shows
an overall positive correlation between abundance and potential activity. However for
certain populations these two measurements are uncoupled: whereas some rare
populations like Alteromonas demonstrate high potential activities, some dominant
populations such as SAR11 are abundant but less active. The cyanobacterium
Prochlorococcus exhibits disproportionally higher activity than their biomass indicates,
and their vertical activity profiles are consistent with RT-qPCR assessments. Our results
suggest that abundance and activity are conceptually and quantitatively different, and
each provides the other new dimension of information, which as a combination will help to better understand the community dynamics and ecology. However, factors such as cell size, rDNA operon numbers, and species specific ribosome efficiency may affect the comparison between populations, and the sequencing library based rRNA: rDNA ratio can be skewed by variation of community compositions. Furthermore, to quantitatively interpret the field rRNA data as ecological meaningful activities (such as growth rate), theoretical/empirical functions are still lacking. Key environmental populations of particular functional interests could be identified using this rRNA/rDNA sequencing technique; then the individual relationship between rRNA and specific activity could be resolved in lab, and eventually applied back to the field to interpret rRNA patterns. On the other hand, the power of this rRNA/rDNA ratio method lies in comparison with associated environmental variables, based on which sensitive activity changes of certain community members could be captured and related to environmental drivers. To achieve this, more community rRNA/rDNA sequencing needs to be conducted across spatial or temporal (Campbell and Kirchman 2012) environmental variations.

As a tool to assess the clade specific activity, the methods we developed can be used to answer more activity related questions or test ecological hypothesis. For example, Jiao et al. have discovered “deep Prochlorococcus” signal with up to ~10^3 cells ml^{-1} in aphotic zone (300 – 800 m) based on flow cytometry counts. One hypothesis is that, vertical mixing processes could transport cells back and forth between euphotic
and aphotic zones, allowing them to access temporally and spatially alternating light and nutrient conditions, which could support higher production in the euphotic zone and longer survival in the aphotic zones. To address this hypothesis, the cells have to be proved viable first. Based on the RT-qPCR method, we identified Prochlorococcus cells at 300 m maintaining significantly amount of cellular rRNA (Fig. 17), which indicates potential viability. And interestingly the detected rRNA belongs to a high-light adapted ecotype eHL-II, probably an adaption to frequent vertical mixing processes within the mixed layer. After normalized by SSC, the rRNA cell\(^{-1}\) SSC\(^{-1}\) value is close to the intercept of our rRNA - growth standard curve (Chapter 2). This suggests that those deep Prochlorococcus may minimize their metabolism to survive through extended darkness.

Moreover, a combination of independent method based estimations of different activity metrics is more robust and informative than any individual technique. As shown rRNA is a suitable activity indicator for mixed natural populations, and besides RT-qPCR technique other detection methods have been developed such as FISH and CARD-FISH (catalyzed reporter deposition-fluorescence in situ hybridization) (Pernthaler et al. 2002). Another promising direction is flow cytometry sorting coupled with physiology dyes or radioactive tracers, e.g. cell sorting based ETS activity (Del Giorgio et al. 1997), primary production (Jardillier et al. 2010; Li 1994), and nutrients uptake rates (Casey et al. 2007). Knowing each approach has its strength and
limitations, we can improve the robustness by multi-level assessments and in the meanwhile expand current environmental applications to yield more ecologically meaningful and statistically reliable data as well as compare with abundance and diversity patterns; all this will help to move the field forward.
Figure 17: Vertical profiles of (A) light (% surface PAR) and temperature (°C), (B) eHL-II (eMIT9312), eLLIV (eMIT9313) and total Prochlorococcus abundance (cells ml⁻¹), (C) rRNA cell⁻¹ for eHLII and eLLIV, and (D) rRNA cell⁻¹ normalized by SSC. Arrows point to potentially viable deep Prochlorococcus.
Appendix

DNA extraction and qPCR (Figure 18A): The DNA yield from the default extraction method (i.e. DNA preservation buffer and 15 min heat lysis) is among the average yields of the treatments without any RNA preservation solution. All three types of RNA preservation solution, including RLT with 1% 2-mercaptoethanol, RNA Shield and RNAlater, inhibit qPCR reaction even after 100-fold dilution. And there is no significant difference caused by bead-beating, or 15 min heat lysis, or DNA preservation buffer.

RNA extraction and RT-qPCR (Figure 18B): The highest yield is from samples preserved in RLT with 1% 2-mercaptoethanol (default treatment), and the second is in RNA Shield. Other preservation methods have much lower efficiency, i.e. less than 50% RNA yield of the top two. And bead-beating step does not make significant difference.
Figure 18: Preservation and lysis methods comparison. A) DNA extraction and 23S gene qPCR. B) RNA extraction and 23S gene RT-qPCR. Three RNA preservation solutions are RLT (Qiagen) with 1% 2-mercaptoethanol, RNA Shield (Zymo Research) and RNAlater (Ambion). Lighter colored bars are treated with bead-beating. Arrow points to the default RNA or DNA extraction method. Asterisks indicate values calculated from 100-fold diluted samples which cannot be amplified at original concentration due to qPCR inhibition.
References


Chisholm, S. W. and others 1992. *Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b*. Archives of Microbiology 157: 297-300.


Lin, Y. and others 2013. In situ activity of a dominant Prochlorococcus ecotype (eHL-II) from rRNA content and cell size. Environmental Microbiology.


121


Williams, T. J. and others 2012. A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. The ISME Journal.


populations in oligotrophic environments. Aquatic Microbial Ecology 30: 159-174.


Biography

Yajuan Lin was born on July 10th, 1984, in Chongqing, P. R. China to Wenyun Lin and Liyuan Zheng. She was raised by Yangtze River, and its water will always be in her blood. She graduated from Peking University on June, 2006, with a Bachelor’s of Science double majored in Geology and Biology. During her Ph.D. training in Biological Oceanography, including 2.5 years at University of Hawaii and 3.5 years at Duke, she authored and coauthored the following papers:


