THE ECOLOGY AND EVOLUTION OF THERMOPHILIC *SYNECHOCOCCUS* SPECIES

by

Eric Daniel Becraft

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Ecology and Environmental Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

January 2014

APPROVAL

of a dissertation submitted by

Eric Daniel Becraft

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency and is ready for submission to The Graduate School.

Dr. David M. Ward

Approved for the Department of Land Resources and Environmental Sciences

Dr. Tracy Sterling

Approved for The Graduate School

Dr. Karlene A. Hoo

DEDICATION

To my mom, who couldn’t see me make this day.

To my dad, who helped me all along the way.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation Frontiers in Integrative Biology Research Program (EF-0328698), the National Aeronautics and Space Administration Exobiology Program (NNX09AM87G), the Danish Council for Independent Research | Natural Sciences (to Michael Kühl and Sheila Jensen), and the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER′s Genomic Science Program (GSP) 395. This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL) under contract 112443. This study was conducted under Yellowstone National Park research permits YELL-0129 and 5494 (David Ward) and YELL-02058 (Michael Kühl), and we appreciate the assistance from National Park Service personnel. I would also like to thank the following people for their support during the completion of this project: Melanie Melendrez, Mary Bateson, Chris Klatt, Jason Wood, Shane Nowack, Isaac Klapper, Millie Olsen and Jennifer Weeding. I would like to individually thank my mentor, Dr. David M. Ward, for helping me become a better scientist throughout my graduate studies.

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

TABLE OF CONTENTS – CONTINUED

TABLE OF CONTENTS – CONTINUED

TABLE OF CONTENTS – CONTINUED

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

Table Page

LIST OF TABLES - CONTINUED

Table Page

LIST OF TABLES - CONTINUED

Table Page

LIST OF TABLES - CONTINUED

Table Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES - CONTINUED

Figure Page

LIST OF FIGURES - CONTINUED

Figure Page

LIST OF FIGURES - CONTINUED

Figure Page

LIST OF FIGURES - CONTINUED

Figure Page

LIST OF FIGURES – CONTINUED

Figure Page

LIST OF FIGURES - CONTINUED

Figure Page

LIST OF FIGURES - CONTINUED

Figure Page

MOST COMMONILY USED TERMS AND ACRONYMS

BAC:	Bacterial artificial chromosome
DGGE:	Denaturing gradient gel electrophoresis
DV:	Dominant variant
Ecotone:	A transitional zone between two ecological communities, as between a forest and grassland or a river and its estuary
Ecotype:	A population of microorganisms that occupy a distinct ecological niche, equivalent to ecological species.
EF:	Ecotype formation
ES:	Ecotype Simulation; Simulation of the evolutionary history of organisms sampled from nature according to the Stable Ecotype Model.
EZ:	Ecological zone; defined area between ecotones.
HFS:	High-frequency sequences; ≥50 sequences in a set of samples.
ITS:	Internal transcribed spacer region separating the 16S rRNA and 23S rRNA loci.
LFS:	Low frequency sequences; <50 sequences in a set of samples.
MLSA:	Multi-locus sequence analysis
MS:	Mushroom Spring
OS:	Octopus Spring
PE:	Putative ecotype
PS:	Periodic selection
SLA:	Single locus analysis

ABSTRACT

To understand the ecology of any environment, the fundamental species-like units that interact with the biotic and abiotic components of that environment must be understood. Previous research conducted in the Ward Lab has shown that 16S rRNA genotypes have unique distributions along the effluent flow path, and that 16S-23S rRNA internal transcribed spacer region genotypes had unique distributions along vertical gradients at some temperatures. However, out of concern that these genetic markers were too conserved to accurately detect ecological species, in this dissertation I analyzed *Synechococcus* cyanobacterial diversity in Yellowstone National Park using the more highly-resolving *psaA* (Photosystem I reaction center protein) locus and an evolutionary simulation based on the Stable Ecotype Model to demarcate putative ecotype populations that are hypothesized to be ecologically distinct species. I used denaturing gradient gel electrophoresis to examine fine-scale distributions of predicted ecotypes along flow and vertical gradients. Because this approach was not sequence-based, and was limited in depth of sequence sampling and habitat coverage, I employed next-generation sequencing technologies (e.g. Ti454-barcoding and sequencing), which was based on sequence information and allowed much deeper sampling and habitat coverage. I used Ti454-barcoding to extend fine-scale distribution studies, examine temporal differences in ecotype-specific gene expression, population-specific responses to environmental perturbations and to examine within-ecotype population genetics. Additionally, I used Ti454-barcoding to analyze frozen samples collected in previous Ward Lab studies, which enabled examination of (i) the biogeographic distributions of *Synechococcus* sequence variants across the Northwestern United States, (ii) the long-term stability of ecological populations in Octopus Spring and Mushroom Spring, and (iii) the initial colonization of disturbed mats. I was also able to use this technology to contribute to other ongoing Ward Lab studies, including (iv) testing the purity of *Synechococcus* isolate cultures before genome sequencing, (v) distribution analyses of loci used in previous population genetics studies, and (vi) linking *psaA* distributions to ecotype-specific genomes. I was able to confirm that many abundant *Synechococcus* putative ecotypes predicted by Ecotype Simulation are ecologically distinct populations, containing ecologically homogeneous individuals. The conclusion of my research supports applying the ecological species concept to hot spring *Synechococcus* cyanobacterial populations.

CHAPTER 1

INTRODUCTION

The debate on what constitutes a microbial species has become more complex with the advent of modern molecular technologies. The immense genetic diversity in nature, along with the discovery of the role genetic exchange has had on the evolution of microorganisms throughout Earth's history, have caused many theories on microbial species to surface, including the hypothesis that microbial species might not exist at all (Doolittle and Papke 2006; Papke et al., 2007). However, many studies have shown that populations of microorganisms can be differentiated by their distinct ecological properties (Ferris et al., 1996a; West and Scanlan, 1999; Ramsing et al., 2000; Rocap et al., 2002; Cohan et al., 2006; Johnson et al., 2006; Ward et al., 2006; Sabehi et al., 2007; Connor et al., 2010; Denef et al., 2010). Microbes exist in almost all known environments on Earth, and have many unique life histories that expose them to different combinations of forces causing or acting upon genetic variation, such that a single species concept that would encompass all microorganisms is unlikely (Ward et al., 2008). However, the identification of the species-like units is paramount to understanding the ecology and evolution within microbial communities.

Species Concepts for Eukaryotic Organisms

While there is still debate as to what constitutes a eukaryotic species (Mayden, 1997), boundaries between populations are more easily observed. In the Biological

Species Concept, species consist of populations of individuals that are capable of reproducing with one another and are reproductively isolated from other organisms (Mayr, 1942, 1982). The Biological Species Concept was later expanded to include the ability to produce fertile offspring and mate recognition strategies (de Queiroz, 2005, 2007). This reproductive isolation definition is widely used for eukaryotic organisms, but is difficult to apply to microbial populations. Microorganisms present unique challenges when one is attempting to organize groups of individuals into species-like populations because they are asexual, even though they can exchange DNA with other unrelated microorganisms through a variety of genetic exchange mechanisms.

In the Evolutionary Species Concept species have unique evolutionary roles and/or fates separate from other populations of organisms (Simpson, 1961; Mayden, 1997). While this definition can also encompass microbial species, knowing when divergent populations should be judged to be evolutionarily distinct is problematic. The Ecological Species Concept expanded on the definition of evolutionary species to include the ecology of the organisms (Van Valen, 1976) by proposing that individuals within the same species must share the same ecological niche, allowing a less arbitrary view of what constitutes a microbial species.

What are Microbial Species?

In an attempt to organize the vast diversity contained in the prokaryotic domains of life, multiple techniques have been developed to identify and classify microbial species. Traditionally, phenotypic characteristics were used to differentiate

microorganisms. More recently, with the advent of modern molecular techniques, genetic methods have become more commonplace. Initially, >70% DNA/DNA hybridization was recommended to demarcate strains within the same species (Wayne et al., 1987). But this was a measure derived from examining genetic differences among strains of species that had already been demarcated on the basis of phenotypic distinctions. Thus, this molecular cutoff considers named species to be real species. The DNA-DNA hybridization cutoff was followed by the use of 2-3% 16S rRNA sequence divergence to demarcate species, which roughly correlated with the >70% DNA/DNA genome hybridization technique (Stackenbrandt and Goebel, 1994; Goodfellow et al., 1997). As such, this cutoff was also based on the demarcations of microbial species that had already been named. As genomic analyses were developed, concerns arose as to whether the molecular resolution in a conserved genetic locus (such as 16S rRNA) was sufficient to accurately demarcate microbial species, which subsequently led to a suggested ≥95% average nucleotide identity (ANI) species demarcation cutoff (Konstantinidis and Tiedje, 2007; Konstantinidis, 2011).

Studies of microbial genetic diversity in nature highlighted complications with the traditional approaches of cultivation and microscopy. First, through the examination of multiple environments it was realized that organisms cultivated within the laboratory are usually not those that are predominant in the natural environment (Giovannoni et al., 1990; Ward et al., 1990; Amann et al., 1995; Hugenholtz et al., 1998; Head et al., 1998). Second, studies that examined the molecular diversity of various environments showed that systems that were thought to be simple are actually complex communities consisting

of morphologically similar taxa, some containing many ecologically distinct, but closely related populations (Ferris and Ward, 1997; Ferris et al., 2003), and complex ecosystems such as soils and marine environments can consist of millions of species of microorganisms (Dykhuizen, 1998 analysis of the data of Torsvik, 1990). Lastly, while phylogenetic analysis of the universally conserved 16S/18S rRNA molecule allowed for the mapping of the tree of life (Woese, 1987; Woese et al. 1990), including the discovery of the Archeal Domain of life, 16S rRNA divergence cutoffs tend to lump ecologically distinct microbial populations together because of the highly-conserved nature of this locus. The "operational" taxonomic units demarcated as species using such molecular cutoff values might actually represent many ecological species, each adapted to its local environment. If true, the ecological diversity is underestimated because it is being examined through an anthropocentric lens that applies molecular cutoffs that essentially demarcate species as what microbiologists previously defined those species to be based on phenotypic properties.

Microbiologists have different ideas about what constitutes a species, and how species relate to the principles of ecology and evolution (Gevers et al., 2005; Cohan and Perry, 2007). A laissez-faire point of view supports that the traditional method of defining microbial species by their morphology and phenotypic traits is a practical way to demarcate species (Roselló-Mora and Amann, 2001). While this approach can be used in the identification of well-studied pathogens, it ignores the ecology and evolution of the organism, and vastly underestimates the diversity that exists in nature. Another, extreme viewpoint is that microbial species do not exist (Doolittle and Papke, 2006, Papke et al.,

2007). For instance, Doolittle and Papke (2006) argue that lateral gene transfer, within-species genomic variability and homologous recombination make it hard to imagine a single species concept for all genetic coherence. These authors state that, "*Whether or not bacteria have species is a perennially vexatious question. Given what we now know about variation among bacterial genomes, we argue that there is no intrinsic reason why the processes driving diversification and adaptation must produce groups of individuals sufficiently coherent in their genetic and phenotypic properties to merit the designation 'species' -- although sometimes they might.*" Over the past few decades sequenced microbial genomes have given evidence of a long evolutionary history of microbial recombination and viral transduction (Xavier et al., 2010). This has caused some to hypothesize that lateral gene transfer is rampant so that the existence of microbial species is implausible in most cases (Boucher et al., 2003). However, gene content differences between close relatives do not necessarily indicate ecological distinctness (Touchon et al., 2009). Lastly, there are those that believe that the basic principles of ecology and evolution developed by the examination of eukaryotes hold true for microorganisms as well (Ward, 1998; Cohan 2002; Ward and Cohan, 2005, Ward, 2006). In the latter case, microorganisms evolve and adapt to unique environmental niches or though geographic isolation generating evolutionarily distinct lineages that can be observed by examining genetic diversity in the environment. Recombination is an important factor in the evolution of microorganisms, but while they can be the catalysts of speciation events, the rates of recombination are not high enough to stop the divergence of separate

phylogenetic lineages (Cohan, 2002; Reno et al., 2009; Vos and Didelot, 2009; Melendrez et al., in preparation).

Ward Lab Studies of Mushroom Spring and Octopus Spring Microbial Mat Communities

The Ward Lab has been researching the microbial mats of Mushroom Spring (Figure 1.1) and Octopus Spring (Figure 1.2A) in Yellowstone National Park for over thirty years. Phototrophic mats in alkaline siliceous hot springs in the Upper Geyser Basin (Figure 1.2B) can be millimeters to centimeters thick. The ~1 mm-thick top green layer of the microbial mats at Mushroom Spring and Octopus Spring is dominated by sausage-shaped cyanobacteria assigned to the named genus *Synechococcus*, which occur in population densities as high as 10^{10} cells per cm^3 (Bauld and Brock, 1973; Brock, 1978; Ward et al., 1989; Ward et al., 1998) (Figure 1.2 C and D). These *Synechococcus*

Figure 1.1. Mushroom Spring showing cyanobacterial mat in foreground and spring in background. (Reprinted from Ward et al., 2006).

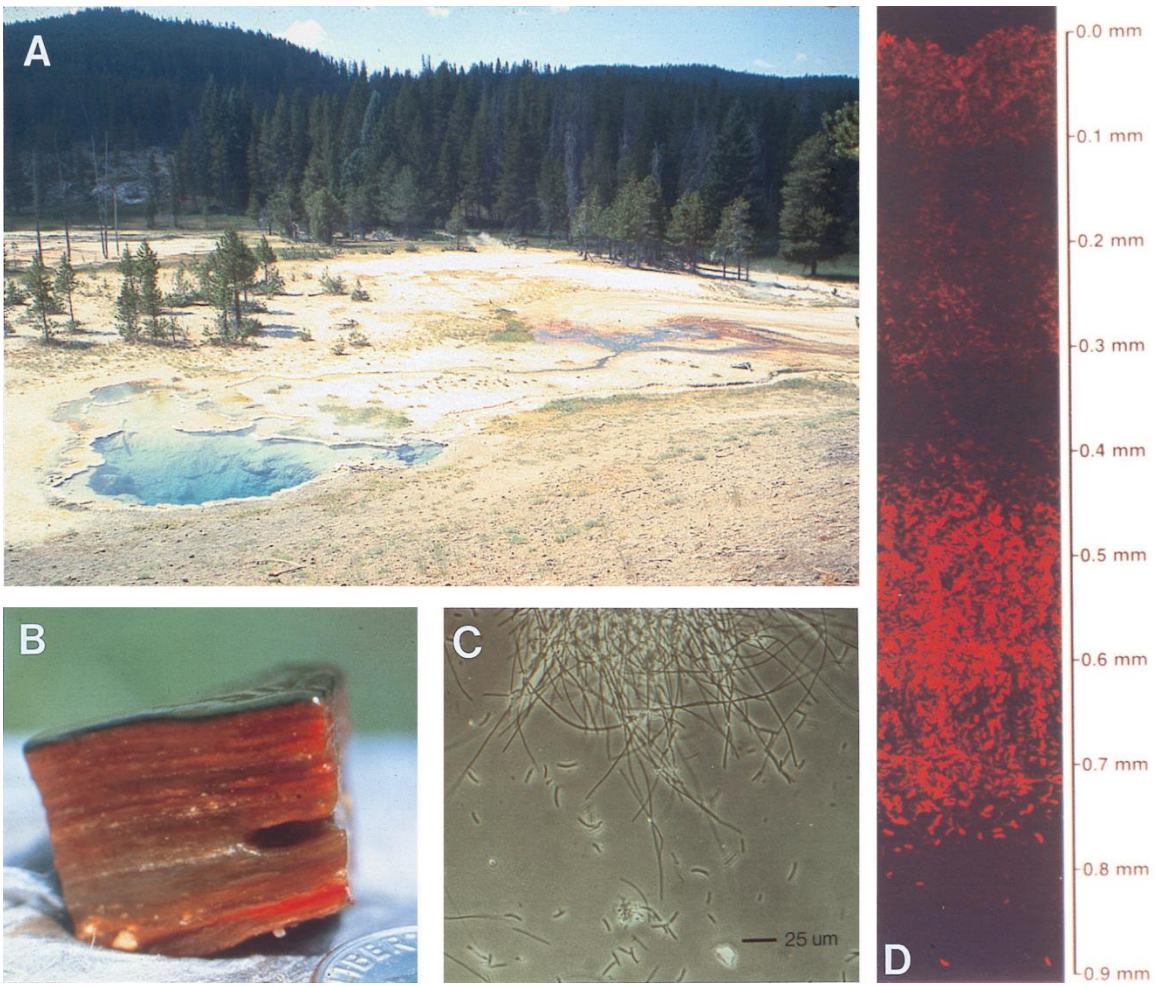


Figure 1.2. Laminated cyanobacterial mat community of Octopus Spring, Lower Geyser Basin, Yellowstone National Park viewed at different scales. (A) Landscape image showing green-orange mat in the effluent channel. (B) Cross section of a cyanobacterial mat sample from ~60°C. (C) Phase-contrast microscopy image of homogenized 1-mm thick upper green layer of mat showing the predominant cyanobacteria, sausage-shaped *Synechococcus* populations, embedded in a matrix of filaments. (D) Autofluorescence microscopy image of a horizontal cryotome section through the upper 1 mm-thick green layer of the 61°C Mushroom Spring mat showing banding of *Synechococcus* populations at different depths (Reprinted from Ward et al., 1998).

spp. share the upper mat layers with filamentous anoxygenic phototrophs assigned to the named genera *Roseiflexus* and *Chloroflexus*, as well as an *Anaerolineae*-like population of *Chloroflexi* (Figure 1.2C), other anoxygenic photoheterotrophic bacteria, including *Candidatus* Chloroacidobacterium thermophilum and *Chlorobiales*-like organisms, and

novel chemoorganotrophic populations (Nübel et al., 2002; Klatt et al., 2011). Cyanobacterial mats in Yellowstone formed by *Synechococcus* occur between ~45^{o}C and 73^{o}C, the upper temperature limit of photosynthesis (Brock and Brock, 1968; Ferris et al., 1996a; Ward et al., 1998; Miller et al., 2000). Studies of species in the Ward Lab have focused on the *Synechococcus* populations because of the seemingly homogenous distribution of these unicellular cyanobacteria along both flow and vertical gradients.

Comparison of Cultivation and 16S rRNA Approaches

The most readily cultivable *Synechococcus* strain (C1, which is closely related to *S. lividus* and *Thermosynechococcus elongataus*) from Mushroom Spring and Octopus Spring is not the most abundant cyanobacterium in these mats as once thought (Meeks and Castenholz, 1971; Ferris et al., 1996b). Molecular analysis based on 16S rRNA cloning and sequencing identified that a group of closely related *Synechococcus* referred to as A-like and B-like populations are the predominant cyanobacterial members of the community (Ward et al., 1990; Ferris et al., 1996b; Ward et al., 1998). Five closely related, but different cyanobacterial 16S rRNA genotypes were detected, which were labeled as A″, A′, A, B′, and B (Ferris and Ward, 1997) (Figure 1.3), none of which were phylogenetically similar to *Synechococcus lividus* (OS C1 Isolate in Figure 1.3).

Distribution of Molecular Variants

Denaturing gradient gel electrophoresis (DGGE) has been used to examine genetic diversity in marine and terrestrial habitats (Muyzer et al., 1993; Ferris et al.,

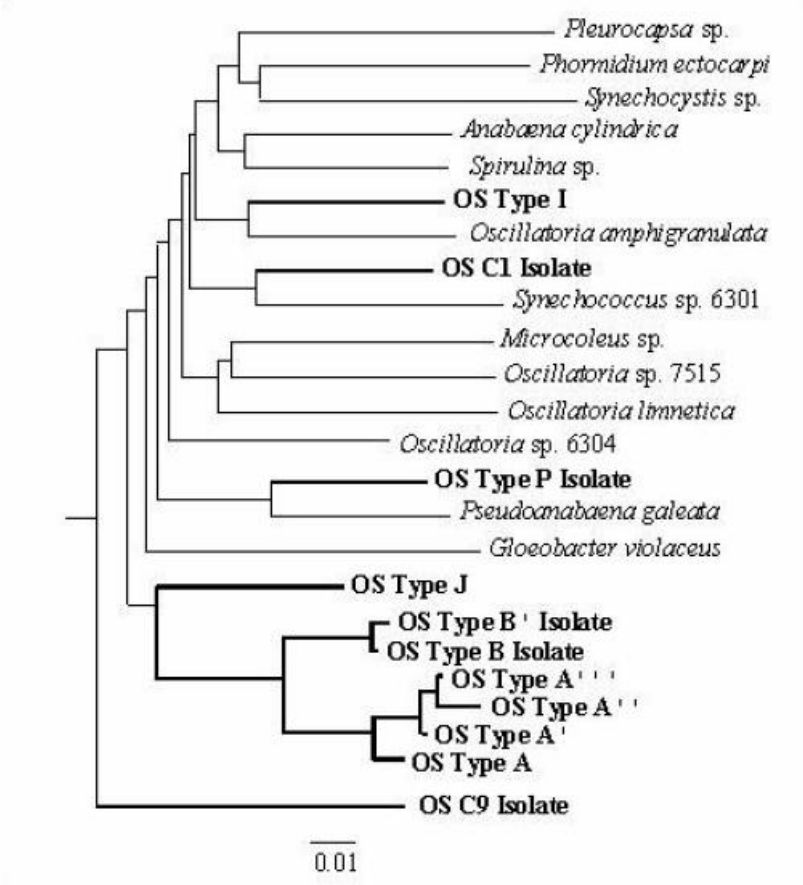


Figure 1.3. Distance matrix phylogenetic tree depicting cyanobacterial 16S rRNA sequences, which form a monophyletic lineage in Domain Bacteria, including those from the Octopus Spring mat. The scale bar corresponds to 0.01 nucleotide substitutions per sequence site. *Synechococcus lividus* is represented by the OS C1 isolate. The cyanobacterial sequences detected in the Octopus Spring mat are highlighted in bold. Genotype OS Type A‴ was detected after a disturbance experiment (Ferris et al., 1997). (Reprinted from Ward et al., 1998)

1996a). DGGE uses a gradient of increasing denaturant concentrations to melt apart double-stranded DNA based on sequence differences. Primers used for PCR amplification have an end clamp sequence containing ~30 guanosine and cytosine residues. These nucleotides form base-pairs with more hydrogen bonds than adenine and thymine. Consequently, when the portion of the double-stranded sequence being amplified reaches a point in the denaturant gradient causing it to melt into single strands,

the end sequence resists melting. This results in a molecular configuration that traps partially melted sequences at different positions in the gel according to differences in their nucleotide sequences. Each unique sequence forms a band in the gel, theoretically migrating to a unique position dependent upon base pair composition and sequence. This allows for the physical separation of unique sequences within an environmental sample, and along ecological gradients, even if only a few base pair differences differentiate genetic variants.

When samples collected along the effluent flow path from the upper temperature limit to ~48°C were analyzed with DGGE, the *Synechococcus* 16S rRNA genotypes were shown to have unique temperature distributions in both Octopus Spring and Mushroom Spring (Figures 1.4 and 1.5, respectively). The A″ and A′ genotypes were found at the highest temperatures from ~72-66^{o}C, the A genotype was found at temperatures from ~67-57^{o}C and the B and B′ genotypes were found at temperatures between ~63 and 48^{o}C (Ferris and Ward, 1997; Ward et al., 2006). *Synechococcus* strains with 16S rRNA sequences corresponding to some of these environmental 16S rRNA genotypes were cultivated and shown to have unique optimal growth temperatures, as well as different temperature ranges, that correlated with their *in situ* distributions (Allewalt et al., 2006) (Figure 1.6). These results led to a realization that *Synechococcus* may form ecological species, possibly similar to the patterns observed with Darwin's finches on the Galapagos Islands (Darwin, 1859).

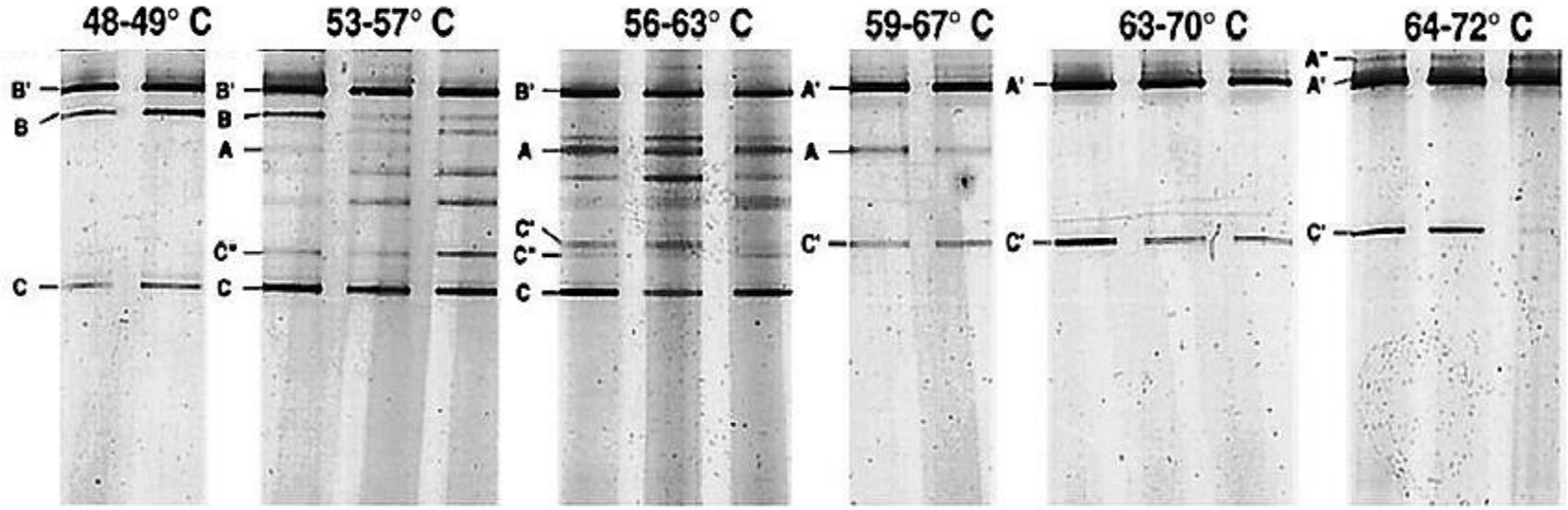


Figure 1.4. Denaturing gradient gel electrophoresis profiles of 16S rRNA gene segments PCR-amplified from DNA extracted from Octopus Spring microbial mat at temperature-defined sites. Letters and primes indicate 16S rRNA genotypes. Genotypes of interest to this dissertation are A, A′ and B′. Multiple lanes under each temperature range represent replicate samples (Modified from Ferris and Ward, 1997).

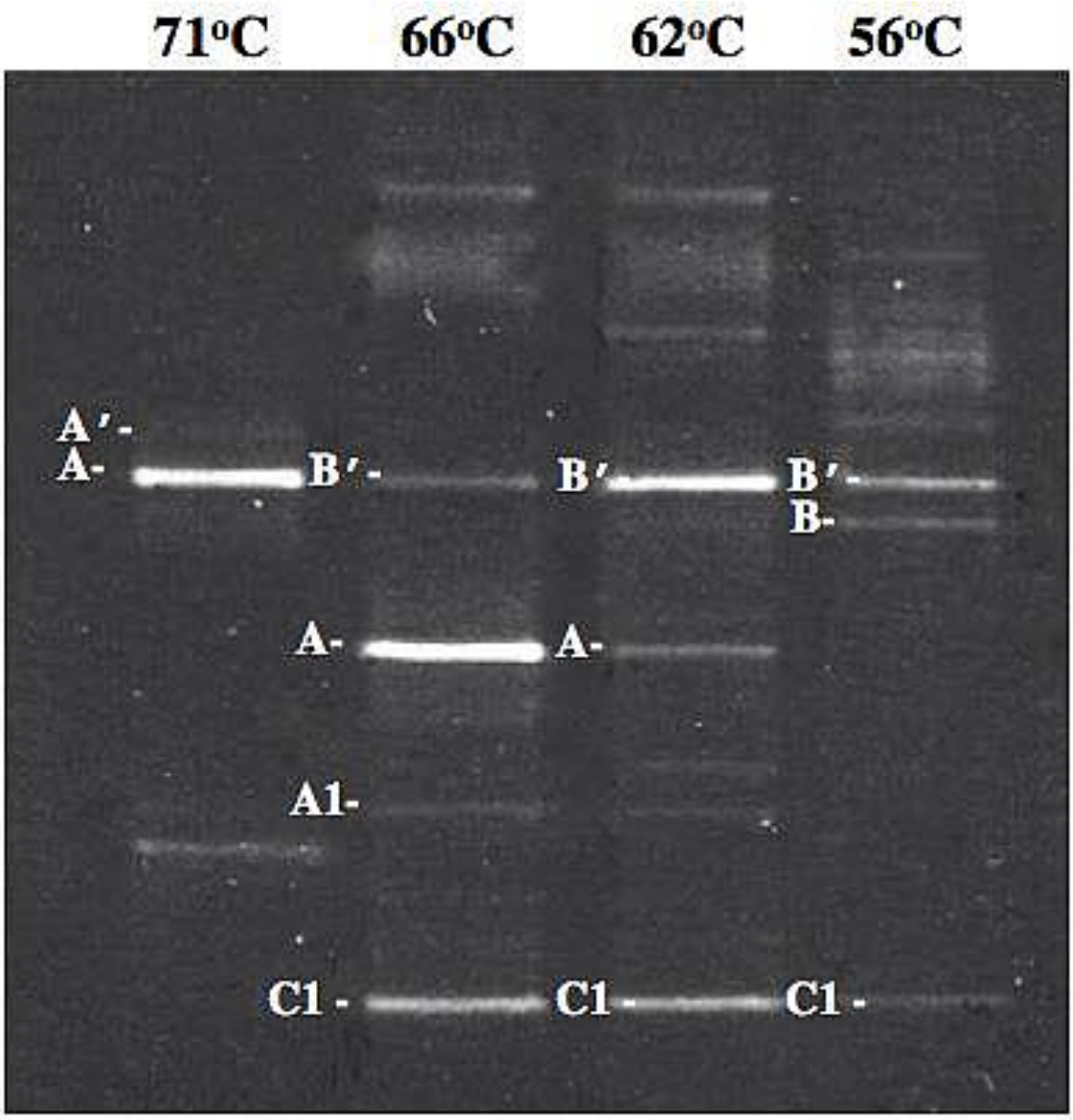


Figure 1.5. Denaturing gradient gel electrophoresis analysis of 16S rRNA gene segments PCR-amplified from mat samples collected at four temperature-defined sites along the Mushroom Spring effluent channel. Letters and primes indicate 16S rRNA genotypes. Genotypes of interest to this dissertation are A, A´ and B′ (Reprinted from Ward et al., 2006).

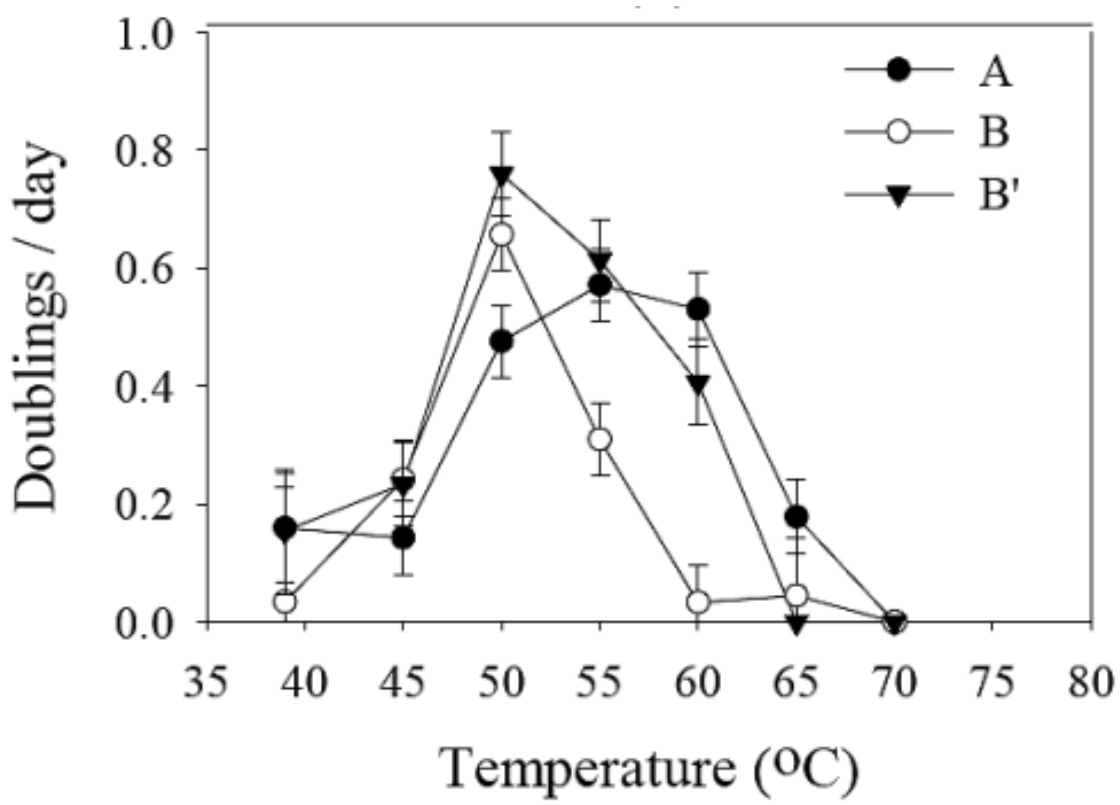


Figure 1.6. Comparison of growth rates of *Synechococcus* isolates with A, B, and B´ 16S rRNA genotypes with respect to temperature. Temperature distributions of *Synechococcus* 16S rRNA genotypes A, B and B′ in the mat were 48-57°C for B, 55-63°C for B′, and 58-65°C for A (refer to Figures 1.4 and 1.5) Error bars correspond to +/- 1 standard error (Modified from Allewalt et al., 2006).

Using a cryotome to dissect mat samples along the vertical aspect, 16S rRNA variation along vertical gradients was also analyzed (Ramsing et al., 2000). At 61°C, the B′ genotype was predominant near the surface of the green layer, while the A genotype was detected at lower depths (Figure 1.7), where a bright band of chlorophyll *a* autofluorescence in *Synechococcus* cells is found (Figures 1.2D). At 68°C, where the A′ genotype is found, there were no genetic differences in the 16S rRNA sequence along the vertical gradient. However, the examination of the more divergent 16S-23S rRNA

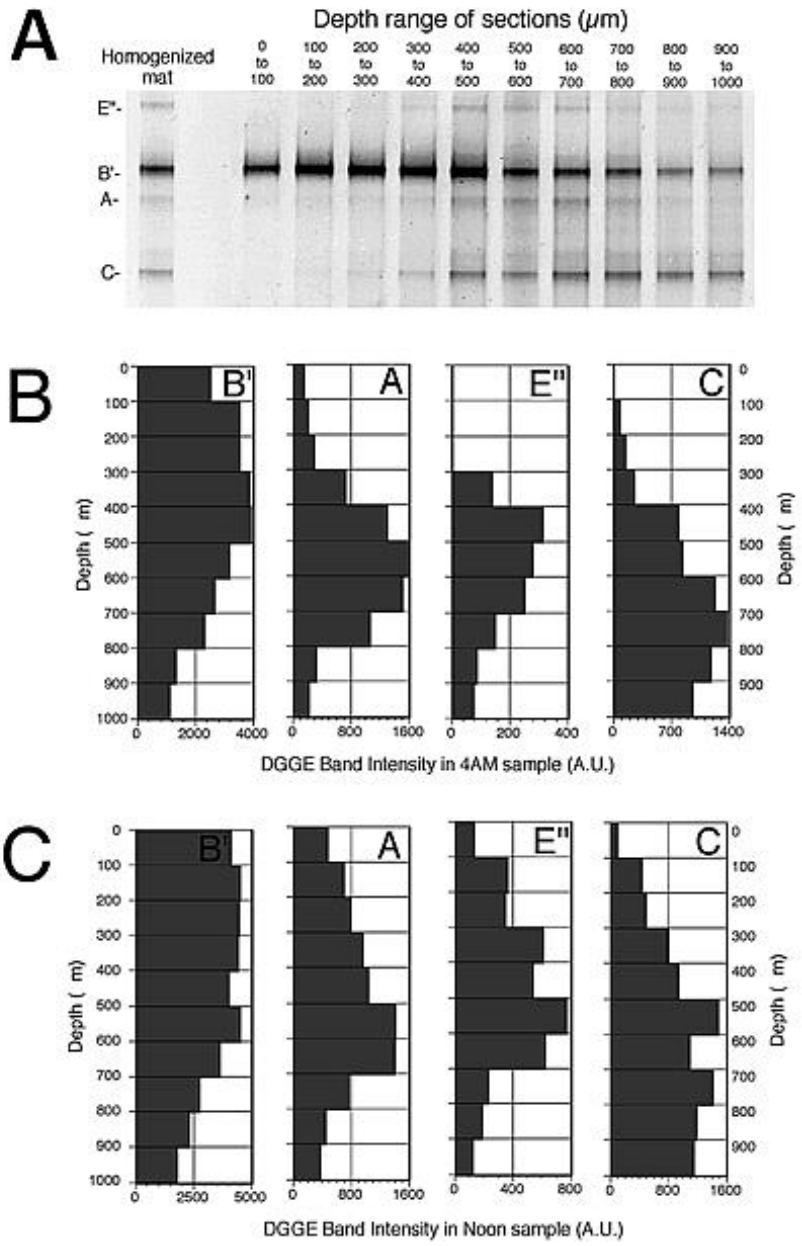


Figure 1.7. Vertical distribution of 16S rRNA-defined genotypes in the 61°C Mushroom Spring microbial mat. (A) Denaturing gradient gel electrophoresis (DGGE) gel of the upper 2-mm cyanobacterial layer and of 100 μm-thick horizontal cryostat sections through the upper 1-mm of the mat's surface layer sampled at 0400 h. Single letters correspond to Octopus Spring 16S rRNA genotypes: A and B′, cyanobacterial; C, *Roseiflexus*-like; E′, green sulfur bacteria-like. (B and C) Bar graphs showing the intensities of individual DGGE bands shown in panel A with depth from samples collected at (B) night (0400 h) and at (C) noon (1200 h). The intensities are depicted in arbitrary units (A.U.) (Modified from Ramsing et al., 2000).

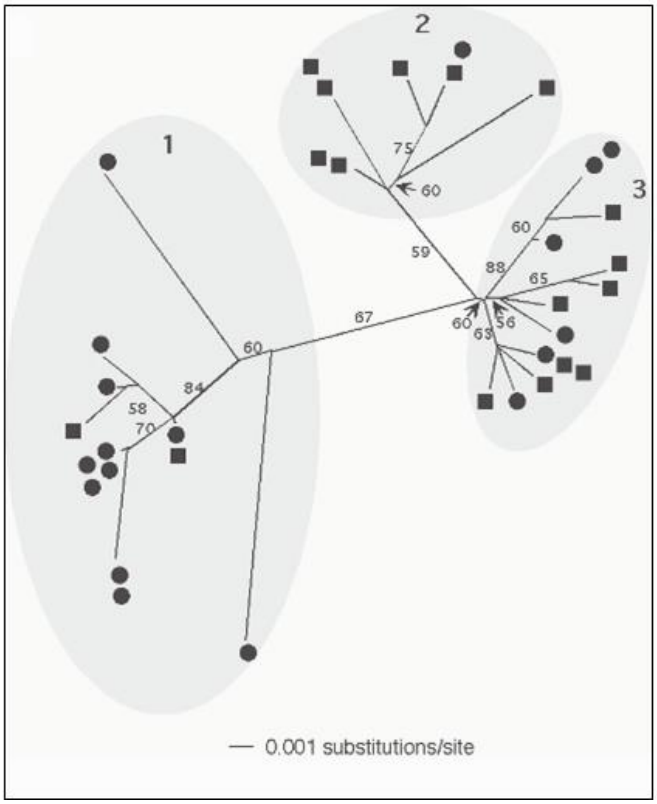


Figure 1.8. Unrooted neighbor-joining tree of thermophilic *Synechococcus* sequences based on 396 nucleotides of the internal transcribed spacer region adjacent to the 16S rRNA gene. Circles are mat-surface sequences; squares are mat-subsurface sequences. Identical sequences are positioned together at terminal branches. Bootstrap values over 50% are shown. Scale bar indicates 0.001 nucleotide substitutions/site (Modified from Ferris et al., 2003).

internal transcribed spacer (ITS) region allowed for the identification of two unique genetic populations that inhabited distinct vertical positions in the mat (Figure 1.8) (Ferris et al., 2003), with the deeper population correlating with the phenotypically distinct band of more intense chlorophyll *a* autofluorescence (Figures 1.2D).

No difference in 16S-23S rRNA ITS sequence was found at 65^{o}C between the top and bottom sections of the green layer despite observed differences in autofluorescing populations with depth, raising the question of whether 16S-23S rRNA ITS had sufficient

molecular resolution to separate more recently evolved ecologically distinct populations (Ferris et al., 2003). Alternatively, the differences in autofluorescence could in this case result from a single species that is differently acclimated to varying light conditions.

This opened up the question of how much molecular resolution was required to detect individuals within species populations (Ward and Cohan, 2005). Additionally, a theory based approach was needed to understand how variation from sampling related to individual genetic variants within ecological species populations as molecular resolution was increased. The ability to understand the evolutionary forces that shape modern-day bacterial lineages, and the ecologies of the fundamental units (i.e. ecological species), is of utmost importance to understanding the structure and function of any microbial community.

Stable Ecotype Theory and Evolutionary Simulation

The Ward Lab began collaboration with Dr. Frederick M. Cohan (Wesleyan University) to apply theoretical modeling to the genetic diversity detected in these mat systems. According to Cohan (2002), species have quintessential properties -- "*Most fundamentally, genetic diversity within a species is thought to be constrained by one or more forces of cohesion" as well as being "irreversibly separate* [from diversity within other species], *with distinct evolutionary fates*", although Cohan (2011) since pointed out that cohesion may not always be required. Bacterial species populations can be held together through periodic selection events, genetic drift or homologous recombination (Koeppel et al., 2008), and can diverge through mutation, geographic isolation, heterologous recombination, niche specialization and genetic drift. These factors can

occur to different degrees in the evolution of different organisms making it difficult to apply a universal microbial species definition to all taxa in all environments (Ward et al., 2008). To accommodate different observations and incorporate the many forces that can act on bacterial diversity, Cohan has developed multiple evolutionary models of bacterial speciation (Cohan and Perry, 2007).

In this dissertation I will primarily focus on the Stable Ecotype Model of species and speciation (Figure 1.9), which best matches patterns of *Synechococcus* diversity in Mushroom Spring and Octopus Spring (Cohan and Perry, 2007; Koeppel et al., 2008; Ward and Cohan, 2005). Species evolving as predicted by the Stable Ecotype Model (i.e. ecological species; which can also be considered ecotypes), "*share many of the dynamic properties attributed to eukaryotic species: genetic diversity within an ecotype is limited by a force of cohesion (in this case, periodic selection); different ecotypes are free to diverge without constraint from one another; and ecotypes are ecologically distinct"* (Cohan, 2002).

In the Stable Ecotype Model, modern phylogenetic diversity is generated through a series of periodic selection and ecotype formation events (PS and EF in Figure 1.9). Bacterial populations accumulate genetic diversity over time though mutation and recombination (dashed fans of diversity in Figure 1.9). When an individual within a population gains a beneficial trait, due to a mutation or recombination event, that individual becomes more fit than other members of the population, thereby purging the genetic diversity within the population though periodic selection (Koch, 1973; Cohan and

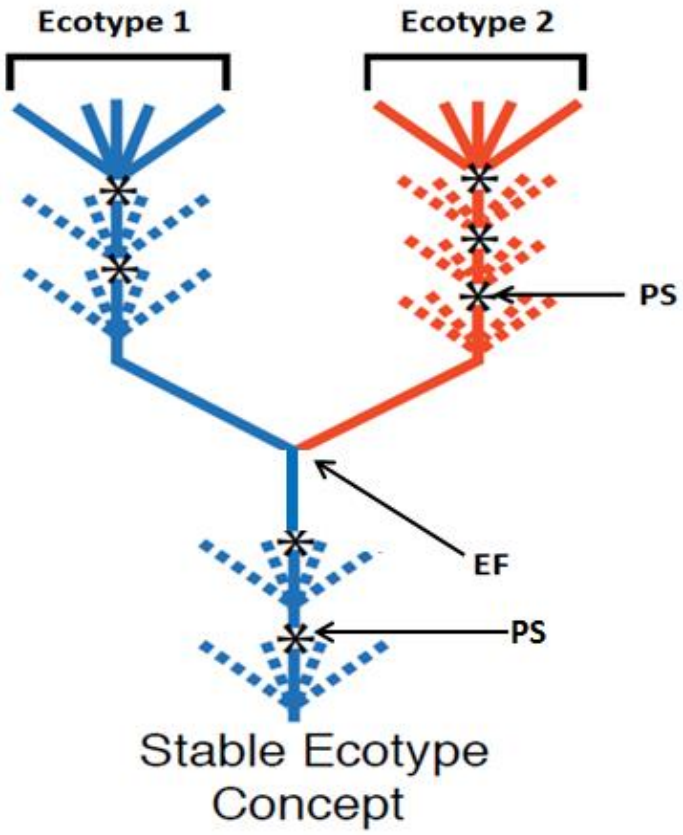


Figure 1.9. The Stable Ecotype Model of microbial speciation depicting the relationship between ecologically distinct populations and DNA sequence clusters. Different ecotypes are represented by different colors. EF shows an ecotype formation event. PS (asterisks) show examples of periodic selection events that purge population diversity except for one most-fit individual. Dashed lines represent extinct lineages (Modified from Cohan and Perry, 2007).

Koeppel, 2008). Hence, a bacterial species can be considered a population of individuals that accumulate mutations within the same ecological niche, and which are held together by occasional collapses in genetic diversity due to a series of periodic selection events. However, a beneficial mutation or recombination event might allow an individual within a population to utilize a new environmental niche (i.e. ecotype formation event), thereby forming a new ecological species that is no longer affected by the periodic selection

events affecting the parental population. The newly formed ecotype then accumulates genetic diversity independently of the parental ecotype and, following a series of periodic selections in each lineage, two different ecotype clades are formed. Genes within populations evolving according to the Stable Ecotype Model should exhibit a 1:1 correspondence between ecotypes and sequence clusters, at least for genes that are not highly recombinant.

Models provide a platform for testing hypotheses in microbial ecology and evolution. Evolutionary simulation programs have been developed to hypothesize which clusters of sequences are ecologically distinct within genetically diverse populations, or which sequences are associated with a specific environment, and thus are a helpful first step in forming hypotheses about the demarcation of species from genetic data. The two models that were used in this study are discussed below.

An evolutionary simulation program called Ecotype Simulation (ES), which is based on the Stable Ecotype Model developed by Koeppel et al., (2008) (Figure 1.10), was used most extensively. In the first step of ES, sequences from nature are binned using different levels of sequence similarity in order to generate a lineage-through-time plot, represented by a curve of increasing number of sequence bins the higher the nucleotide identity used to define a bin (Figure 1.11A). As shown by the left-most vertical bar in Figure 1.11B, when the sequence relatedness criterion is set at 80% sequence divergence, all sequences in the phylogeny shown are members of one bin. However, following vertical bars to the right, as the sequence relatedness criterion is

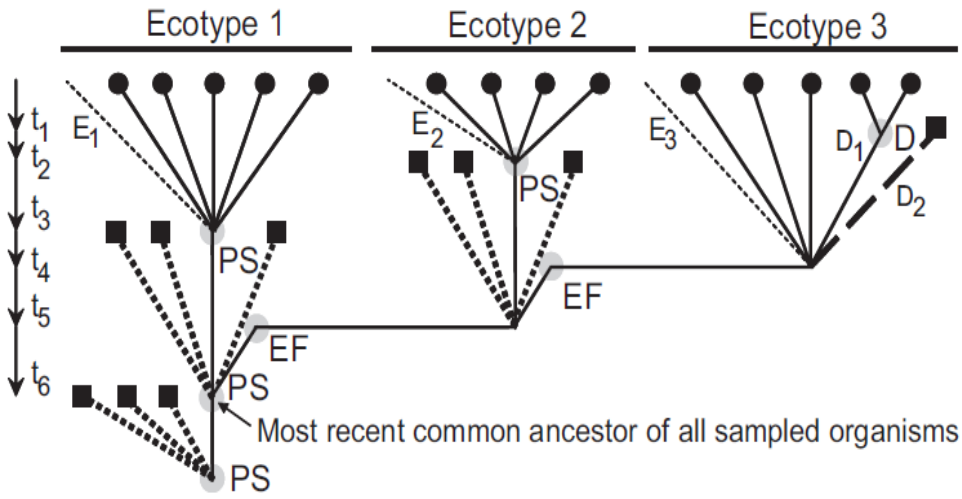


Figure 1.10. The Ecotype Simulation Model for estimating the order and timing (t_n) of periodic selection events (PS), ecotype formation events (EF), the influence of genetic drift (D) and the number of putative ecotypes from sequence data (modified from Cohan and Perry, 2007).

made more stringent, there are more bins each containing fewer sequences. When the sequence relatedness criterion is absolute identity (i.e., 100%), all unique variants are binned separately. The resulting curve (e.g., Figure 1.11A) summarizes the evolutionary history of the genetic diversity in the sample and is used as a reference for the simulation (Koeppel et al., 2008).

In the second step, the ES algorithm generates hypothetical patterns of the evolution of a given set of modern-day variants based on the parameters of the Stable Ecotype Model (Figure 1.10): the number of ecotypes in a phylogeny, the rate of periodic selection events, the rate of ecotype formation events and the impact of genetic drift. ES stochastically generates thousands of possible combinations and timings of these

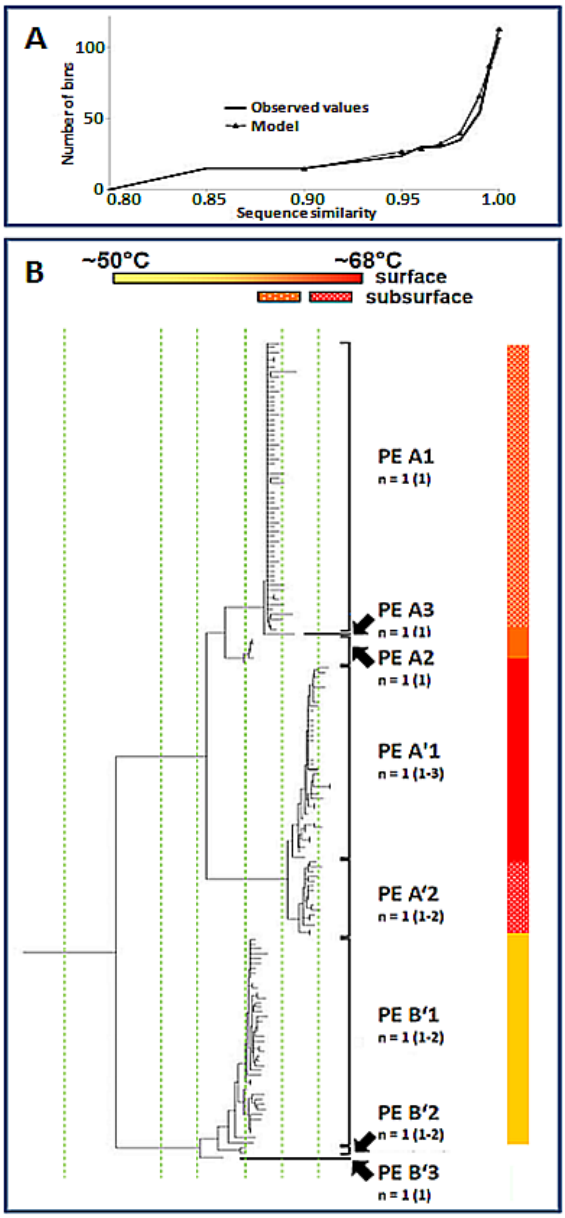


Figure 1.11. Ecotype Simulation analysis of *Synechococcus*-like 16S-23S internal transcribed spacer region sequence data retrieved from the Mushroom Spring and Octopus Spring microbial mats. (A) The actual and modeled relationships between number of sequence bins and the sequence relatedness criterion used to demarcate bins. (B) Phylogenetic tree showing putative ecotypes (PEs) demarcated by Ecotype Simulation, where n is the most likely estimate of the number of PEs and the number in parentheses is the 95% confidence interval. Thin dashed lines in B demonstrate how more bins arise in A as one progresses from the common ancestor at the extreme left of the tree toward individual variants at the extreme right of the tree. Colors highlight location in the mats from which variants predominating a PE clade were collected (see inset key) (modified from Ward et al., 2006).

parameters, each of which coalesces a set of variants in a backward evolution to a common ancestor.

In the third step, ES then runs a forward evolution to simulate the natural history of the phylogeny using each possible combination of the parameters, and the resultant outputs are compared to the reference curve to determine the most likely pattern, and thus the most likely values for each parameter. Finally, ES demarcates clusters of sequences as putative ecotypes (PEs; hypothesized ecological populations) based on the sequences within a clade with the highest likelihood of belonging to a single ecotype (i.e. if the value 1 is included within the 95% confidence interval) (Figure 1.11B). Alternatively, ecotypes can be demarcated using a less-conservative maximum-likelihood approach, which predicts putative ecotypes using the most probable number within the confidence interval (Francisco et al., 2012). ES has been applied to distinct systems such as the *Bacillus subtillus* populations of a west-running canyon in Death Valley that were shown to have unique distributions on cliff faces with varying solar exposure (Connor et al., 2010), and the *Synechococcus* 16S-23S rRNA ITS populations of Mushroom Spring that were shown to have unique distributions along temperature and vertical gradients (Figure 1.11) (Ward et al., 2006).

AdaptML, another program that was used to a lesser extent in this thesis, analyzes genetic diversity from multiple habitats and predicts the number of ecologically distinct populations (Hunt et al., 2008). However, AdaptML focuses on distribution and predicts ecologically distinct populations dependent upon where in the environment sequences were located. It is largely biased by the investigators' ideas about what environmental

features are important (Cohan and Koeppel, 2008). In contrast, ES predicts ecological species without taking environmental features into account, but the environmental data are used to determine whether or not PEs are ecologically distinct (Figure 1.11B). For example, using 16S-23S rRNA ITS data, ES accurately predicted two ecological populations from sequence diversity located at 68°C that were subsequently shown to inhabit surface and subsurface regions of the mat (Figure 1.8 and 1.11B).

High-resolution Molecular Analysis of Protein-Encoding Loci.

Most microbial population genetics analyses to date have been conducted on conserved genetic loci that offer low molecular resolution. Melendrez et al. (2011) analyzed protein-encoding genes with varying degrees of genetic divergence in an attempt to determine the influence of molecular resolution on the sequence-based discovery of ecological diversity (Figure 1.12). The *Synechococcus* loci *apcAB*, *aroA* and *rbsK* were PCR amplified and clone libraries were created separately for the A-like and B′-like 16S rRNA populations. As shown in Figure 1.12, ES predicted more PEs from protein-encoding loci when compared to 16S rRNA plus 16S-23S rRNA ITS (compare Figures 1.11B and 1.12), and as molecular divergence increased, more PEs were subsequently predicted. All 3 loci showed temperature-specific ecotypes at 60^{o}C and 65^{o}C for the A-like populations (Figure 1.12), though only two different temperature-defined habits were used in the analysis, thus limiting the ability to differentiate PEs and prove their ecological distinctness.

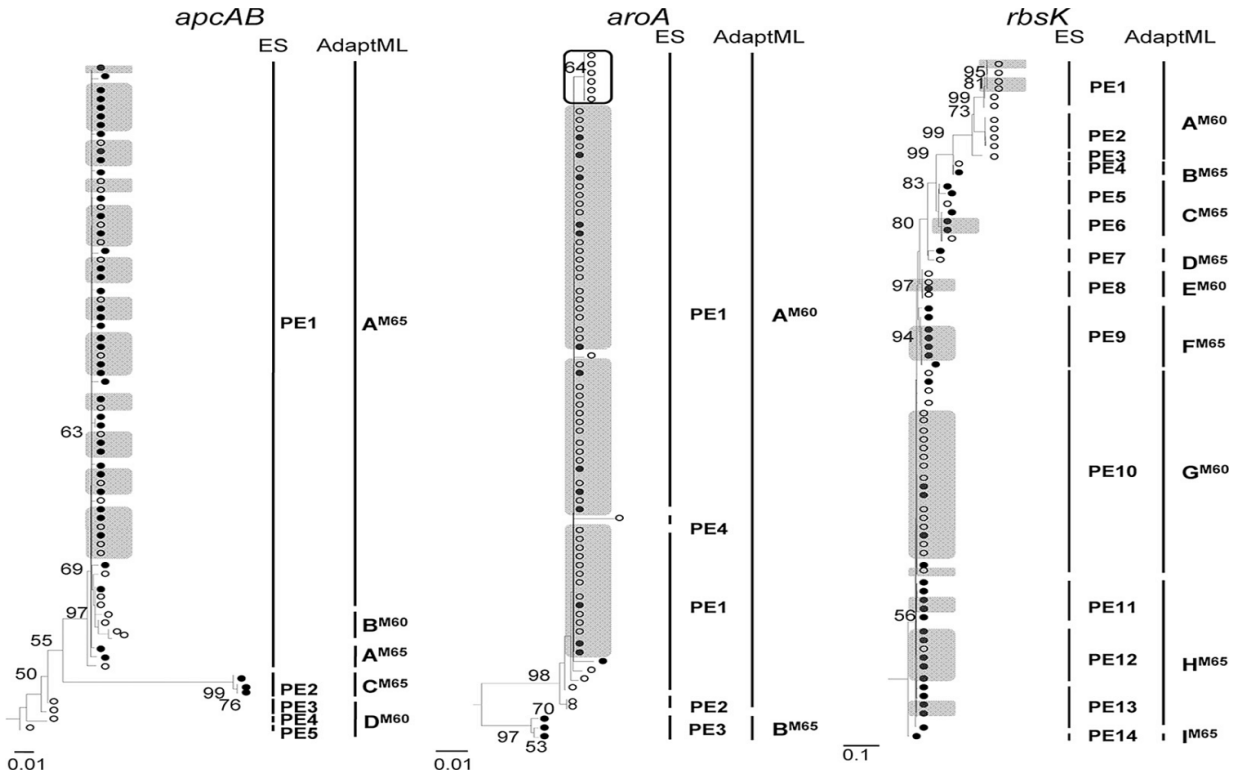


Figure 1.12. Neighbor-joining phylogenies for A-like *Synechococcus* population sequences for *apcA B*, *aroA*, and *rbsK* genes, with putative ecotypes (PE) demarcated by Ecotype Simulation and habitat-specific clusters demarcated by AdaptML (letters). Open and closed circles demarcate sequences from 60°C and 65°C samples, respectively. Replicate sequences of the dominant variant within a demarcated PE are highlighted in shaded boxes. Putative subdominant variants in *aroA* PE1 are highlighted in the unshaded box. Superscripts M60 and M65 next to AdaptML cluster demarcation indicate habitat specificity. Bootstrap values of >50% are shown. Scale bars indicate nucleotide substitutions/site (Modified from Melendrez et al., 2011).

Proving that Predicted Ecotypes are Ecologically Distinct

Identifying ecotypes through population genetics studies and evolutionary simulation programs is only the first step toward analyzing the genetic diversity within the environment. PEs predicted through evolutionary simulation analyses then must be shown to be ecologically distinct (see Figure 1.11B). Previous work was limited by the low molecular resolution of the 16S rRNA locus, minimal sampling of genetic variation and low habitat variability. The aim of this thesis was to evaluate the ecological distinctions of PEs predicted by ES using a more highly resolving protein-encoding locus by analyzing (i) population distributions across a large number of habitats, (ii) species-specific patterns of gene expression and (iii) unique responses to environmental perturbation. Genetic variation in the *psaA* locus was analyzed because (i) it offers more molecular resolution than the 16S rRNA locus, (ii) it is known to be expressed *in situ* (Liu et al, 2012), (iii) it exists as a single copy in the genomes of mat *Synechococcus* isolates (Bhaya et al, 2007), (iv) the protein subunit it encodes (PsaA) is necessary for photosynthesis, and (v) it shows minimal evidence of recombination in the region studied. In Chapters 2 and 3 of this thesis I examine properties of *psaA*-defined *Synechococcus* ecological species and address the following questions.

1) Do putative ecotypes predicted from *psaA* sequences by Evolutionary Simulation and AdaptML exhibit unique distributions along temperature and depth gradients?

Chapter 2 (Becraft et al., 2011) compares and contrasts the PE outputs of the evolutionary simulation program Ecotype Simulation, and the ecological sequence

clusters predicted by AdaptML. DGGE is then used to analyze the distribution of *psaA* variants along fine-scale flow and vertical gradients, thereby increasing habitat variation and enabling determination of whether PEs are spatially distinct.

2) Will the increased sequence and habitat coverage provided by high-throughput sequencing analyses (Ti454-barcoding and sequencing) increase ecotype coverage and improve fine-scale analysis of properties expected of ecological species?

During my thesis work, high-throughput sequencing techniques became available, which permitted deep-coverage, sequenced-based analyses of *psaA* variants and distributions over a large number of habitats. Ti454-barcoding and sequencing allows for greater sequence sampling depth, and allows for the simultaneous analysis of multiple habitats. This is accomplished by ligating a short sequence segment unique to a given sample onto the PCR primer used to amplify sequences from that sample before sequencing. This allows sequences from different samples to be simultaneously analyzed in high-throughput sequencing runs (e.g., using Ti454 technology), resulting in a very large number of sequences that can be sorted to the samples from which they came based on the sample-specific barcode sequence. In Chapter 3 (Becraft et al., submitted), deeper sampling and increased habitat coverage provided by Ti454-barcoding and sequencing allowed me to examine PEs that were insufficiently sampled in Chapter 2, and to better understand their distributions using sequence data instead of DGGE band migration. In addition to extending fine-scale distribution studies reported in Chapter 2, I examined temporal differences in ecotype-specific gene expression and unique responses of predicted *Synechococcus* ecological species populations to environmental perturbations.

Deep sequencing also allowed me to gain insight into the genetic structure of ecological populations, and to address the issue of the ecological interchangeability of individuals predicted to be contained within a PE.

Evaluating *Synechococcus* Species in Broader Ecological Contexts

Having demonstrated the value of Ti454-barcoding, I used this approach to examine a large number of samples that had been collected in previous Ward Lab studies (Chapters 4, 5 and 6) and in collaborations with other Ward Lab students (Chapter 7). The following questions were addressed.

1) Will biogeographical patterning of *psaA* sequence variants provide insight into ecological and evolutionary processes that have affected *Synechococcus* populations in the American Northwest?

In Chapter 4, I examine *psaA* sequence diversity across the Northwestern United States within individual springs in different regions. Biogeographical patterns and distributions were examined and analyzed in relation to geographical distance and hot spring chemical properties to better resolve sequence distributions originally reported in Papke et al. (2003). Variation within and between ecological species (or geological species) was also examined to determine if populations were evolving separately among springs and differently within geographic regions. Additionally, regions and springs were sampled at different ecological settings to observe population distributions in springs with different chemical parameters (e.g., pHs) or along ecological gradients within individual springs (e.g., temperature).

2) How stable are *Synechococcus* ecological populations over time?

In Chapter 5, PE distributions and relative abundances were compared between winter and summer months to determine whether a more highly-resolving protein-encoding locus would resolve seasonal patterns of populations changing that were not seen in a previous study based on 16S rRNA sequences (Ferris and Ward, 1997). I also analyzed *Synechococcus* ecological populations demarcated by ES from *psaA* sequence variation in samples collected from Mushroom Spring and Octopus Spring between 1989 to 2009 in order to observe temporal changes in population abundance and genetic structure of PEs. Additionally, I was able to analyze within-PE diversity over a period of years at both Mushroom Spring and Octopus Spring to observe whether populations changed separately within and between springs.

3) What factors govern the recovery of *Synechococcus* populations after disturbance?

In Chapter 6, I report results of analyses of patterns of recolonization in response to physical disturbance of the mat under different light conditions. This permitted evaluation of the roles of dispersal and ecological adaptation in the recovery process. Since a previous study suggested that physically disturbed sites were colonized by upstream *Synechococcus* populations (Ferris et al., 1997), I also analyzed *Synechococcus* PE distributions in the water flow path, which could be important in influencing recolonization of disturbed mat sites.

Collaborative Studies Employing Ti454-Barcode Technology

In Chapter 7, I address other uses of Ti454-barcode technology I implemented

during my graduate studies that included (i) using deep-sequencing technology to test *Synechococcus* isolate culture purity before genomic analyses, (ii) analyzing the distribution patterns over a wider range of habitats of genes used in other single and multi-locus population genetics studies (Melendrez et al. 2011; Melendrez et al. in prep), and (iii) using *Synechococcus* isolate genomes representative of PEs defined at the *psaA* locus to link *psaA* and multi-locus PE distributions. Background for these studies will be provided in Chapter 7.

Conclusions

This dissertation provides a multitude of evidence, including (i) fine-scale patterns of distribution, (ii) temporally varying ecotype-specific expression patterns, (iii) unique population responses to environmental perturbations, (iv) sequence and population-specific biogeographical distributions (v) PE-specific changes in relative abundance and within-population genetic-variation over time, (vi) initial colonization of newly forming mat after environmental disturbance, and (vii) adaptations of *Synechococcus* isolates, verifying that many putative ecotype *Synechococcus* populations demarcated by ES analysis of *psaA* variation are ecologically and/or geographically distinct species that are subject to the evolutionary and ecological principles that guide thinking about plant and animal species.

CHAPTER TWO

FINE-SCALE DISTRIBUTION PATTERNS OF *SYNECHOCOCCUS* ECOLOGICAL DIVERSITY IN THE MICROBIAL MAT OF MUSHROOM SPRING, YELLOWSTONE NATIONAL PARK

Contributions of Authors

Manuscript in Chapter 2

Author: Eric D. Becraft

Contributions: Designed the experimental study, collected samples, conducted molecular analyses, analyzed output data, and primary writer of the manuscript.

Co-Author: Frederick M. Cohan

Contributions: Mentor and developer of evolutionary simulation program used to demarcate ecological species.

Co-Author: Michael Kühl

Contributions: Microsensor analysis and chemical measurements (e.g. pH and light).

Co-Author: Sheila I. Jensen

Contributions: Microsensor analysis and chemical measurements (e.g. pH and light).

Co-Author: David M. Ward

Contributions: Primary mentor, laboratory provider and co-writer.

Manuscript Information Page

Eric D. Becraft, Frederick M. Cohan, Michael Kühl, Sheila I. Jensen, and David M. Ward

Applied and Environmental Microbiology

___ Prepared for submission to a peer-reviewed journal
___ Officially submitted to a peer-review journal
___ Accepted by a peer-reviewed journal
_x_ Published in a peer-reviewed journal

08/22/11
Volume 70, Issue 6, pp. 7689-7697.

Abstract

Past analyses of sequence diversity in high-resolution protein-encoding genes have identified putative ecological species of unicellular cyanobacteria in the genus *Synechococcus*, which are specialized to 60°C versus 65°C in Mushroom Spring microbial mats. Because these studies were limited to only two habitats, we studied the distribution of *Synechococcus* sequence variants at 1°C intervals along the effluent flow channel, and at 80 μm vertical depth intervals throughout the upper photic layer of the microbial mat. Diversity at the *psaA* locus, which encodes a photosynthetic reaction center protein (PsaA), was sampled by PCR amplification and cloning and sequencing methods at 60°C, 63°C and 65°C sites. The evolutionary simulation programs Ecotype Simulation and AdaptML were used to identify putative ecologically distinct populations (ecotypes). Ecotype Simulation predicted a higher number of putative ecotypes in cases where habitat variation was limited, while AdaptML predicted a higher number of ecologically distinct phylogenetic clades in cases where habitat variation was high. Denaturing gradient gel electrophoresis was used to track the distribution of dominant sequence variants of ecotype populations relative to temperature variation and to O_2, pH and spectral irradiance variation, as measured using microsensors. Different distributions along effluent channel flow and vertical gradients, where temperature, light and O_2 concentrations are known to vary, confirmed the ecological distinctness of putative ecotypes.

Introduction

Surveys of sequence diversity have led to the discovery of ecologically adapted populations in extremely closely related microorganisms in multiple taxa inhabiting many different environments. For instance, molecular diversity in *Prochlorococcus* spp. (Johnson et al., 2006; Martiny et al. 2009; West and Scanlan, 1999) and proteorhodopsin-containing bacteria (Béjà et al., 2001) has been found to vary with depth in the marine water column in relation to variation in light intensity and quality and differences in nutrient availability. Unique *Bacillus* sequence clusters were shown to be associated with desert hillsides having distinct solar exposure patterns and soil textures (Conner et al., 2010; Koeppel et al., 2008). Molecular analyses have also shown closely related ecologically distinct populations in marine *Vibrio* (Hunt et al., 2008), acid mine *Leptospirillum* (Denef et al., 2010) and multiple pathogens including *Legionella* (Cohan et al., 2006), *Escherichia* (Manning et al., 2008, Walk et al., 2007, Walk et al., 2009), *Mycobacterium* (Smith et al., 2006), *Salmonella* (Soyer et al., 2009), and *Yersinia* (Zhou et al., 2004). Finally, genetically distinct populations of hot spring *Synechococcus* have been shown to be associated with different temperature and vertical distributions (Ferris and Ward, 1997; Ramsing et al., 2000; Ward et al., 2006).

In the case of *Synechococcus*, five closely related 16S rRNA genotypes (termed A″, A′, A, B′, and B) exhibited different distributions along the effluent channels of Octopus and Mushroom Springs from the upper temperature limit at 72-74°C to ~50°C (Ferris and Ward, 1997; Ward et al., 2006). Cultivated *Synechococcus* strains representing some of these 16S rRNA genotypes exhibited unique optimal growth

temperatures, as well as different temperature ranges for growth, which correlated with the *in situ* genotype distributions (Allewalt et al., 2003). This provided a means of testing the hypothesis that adaptive evolution, at least in part, can explain the distribution pattern observed *in situ*. Vertical distribution studies gave further evidence for adaptation of *Synechococcus* spp. At 60°C, the *Synechococcus* B′ 16S rRNA genotype was the only cyanobacterial genotype detected near the mat surface, while the *Synechococcus* A 16S rRNA genotype was detected only at lower depths in the approximately 1 mm-thick top green layer of the mat (Ramsing et al., 2000), correlating with a brightly autofluorescent subsurface *Synechococcus* population. At 68°C, where the *Synechococcus* A′ genotype is found, there was no sequence diversity at the 16S rRNA locus throughout the vertical aspect. However, examination of the more rapidly evolving 16S-23S rRNA internal transcribed spacer (ITS) region allowed for the identification of two unique genetic populations that inhabited distinct vertical positions within the top green layer (Ferris et al., 2003). One sequence cluster was found in the top portion of the green layer, while the other was found near the bottom of the green layer, where the more brightly autofluorescent *Synechococcus* population resided. In contrast, at 65°C, no differences in 16S rRNA or ITS sequences were found throughout the vertical aspect of the mat, raising the question of whether more molecular resolution was needed to separate ecologically distinct populations of cyanobacteria, or alternatively, whether one differently acclimated population was merely autofluorescing dissimilarly at different light intensities (Ward et al., 2006).

Recently, genetic variation in protein-encoding loci of *Synechococcus* was analyzed with an evolutionary simulation program based on the Stable Ecotype Model of species and speciation (Melendrez et al. 2011). Ecotype Simulation (ES) was used to predict from sequence variation alone putative ecologically distinct populations (i.e., putative ecotypes (PEs)), which group together sequences that are hypothesized to represent ecologically interchangeable individuals (Cohan and Perry, 2007; Koeppel et al., 2008). The algorithm AdaptML (Hunt et al., 2008), which utilizes both habitat specialization and phylogeny, was also used in this study to identify ecologically distinct phylogenetic clusters of sequences (i.e., ecological clusters (ECs)). These two algorithms have previously been shown to yield similar, though not identical demarcations of ecologically distinct populations (Conner et al., 2010; Melendrez et al., 2011). Previous applications of ES and AdaptML have shown that the greater molecular resolution provided by protein-encoding loci increased the number of *Synechococcus* PEs predicted by 4-14 times that observed with 16S rRNA, and the number of PEs predicted for each gene was a function of the rate of evolution of the gene (Melendrez et al., 2011). Validation of whether PEs meet the expectations of true ecologically specialized populations can be achieved through examination of unique spatial distributions relative to different features of the habitat (Ward et al., 2006). However, in the case of hot spring *Synechococcus* populations, studies done to date using protein-encoding sequences have included only two habitats, and this limited identification of habitat associations.

The aim of this study was to more thoroughly examine the distribution patterns of populations predicted from sequence analyses to be ecologically distinct. To do so, we

examined the distribution of genetic variation of a highly resolving protein-encoding locus along fine-scale environmental gradients as quantified by microsensor analysis (Kühl, 2005). We used variation in the *psaA* locus, which encodes a reaction center protein subunit of photosystem I, an essential component of the cyanobacterial photosynthesis apparatus, to predict PEs. This locus was selected as it is one of the most highly expressed genes detected in metatranscriptomic analyses of these mats (Liu et al., 2011), and thus offers the potential to reveal differences in expression patterns among individual ecotypes, which could be useful for discerning adaptive differences among populations that co-exist spatially. Since PE clades contain dominant sequence variants (Melendrez et al., 2011), we were able to use denaturing gradient gel electrophoresis (DGGE) to track the distributions of PEs predicted by ES and ECs predicted by AdaptML. We show that *Synechococcus* populations predicted to be PEs or ECs do exhibit different distribution patterns as expected of true ecological species.

Materials and Methods

Sampling

This study was focused on the 65°C to 59°C region of the major effluent channel of Mushroom Spring, an alkaline siliceous hot spring in the Lower Geyser Basin, Yellowstone National Park, WY, USA. Samples were collected from sites spread over a distance of ~15 m in the main flow path defined by temperatures measured at the specific time of collection, similar to the sample sites in Appendix A Figure A1. Temperatures at specific sites in Mushroom Spring fluctuate over the diel cycle (Appendix A Figure A2)

and seasonally (S. Nowack and I. Klapper, unpublished results). Microbial mat samples were collected on 20 September 2006 at temperature-defined sites along the flow path using a #2 cork borer (19.6 mm^2). The ~1mm-thick top green layer of the mat containing *Synechococcus* cells was removed with a sterile razor blade and immediately frozen on site in liquid N_2. Samples for vertical distribution analyses, which included the orange undermat beneath the green layer, were collected at midday on 13 and 14 September 2008 with a #2 cork borer and were immediately frozen in isopentane kept cold with liquid N_2. These samples were stored during transit on dry ice in 15 ml Falcon tubes packed with Kimwipes at both ends to prevent physical damage as described by Ramsing et al. (200), and transferred to a -80°C freezer for long term storage.

Vertical Dissection

Frozen cores were embedded in Tissue-Tek O.C.T. compound cryoprotectant (Sakura, Torrance, CA, USA) and sliced parallel to the mat surface at 40 µm intervals using a cryotome (Leica CM 1850, Bellevue, WA, USA) with a fixed sterile diamond knife at -20°C. Two adjacent 40 µm slices were combined for DNA extraction. This procedure was repeated throughout the entire top green layer, which ranged in thickness from 500 to 980 µm.

DNA Extraction and Purification

Nucleic acids were extracted and purified with the fastDNA spin kit (Molecular Biosciences, USA) according to the manufacturer′s instructions. Cells were suspended in

manufacturer′s lysis solution, and lysed by bead-beating in a Fastprep Cell Disrupter (Bio101 Savant Instruments, New York) for 45 sec at a speed setting of 6.5 meters/sec.

PCR Amplification, Cloning,
Sequencing and Phylogenetic Analyses

A 557 bp segment of the *psaA* gene (positions 279-836; total gene length 2268 bp; single copy in the *Synechococcus* spp. strain A and B′ genomes) was chosen for analysis as it resides within the most variable region of the *psaA* gene and exhibits 8% nucleotide divergence between homologs in the *Synechococcus* spp. strain A and B′ genomes (Bhaya et al., 2007). Primers for the amplification of *Synechococcus* A/B-lineage *psaA* genes (psaAforward: 5′-ctgagcggcatgtactacca-3′, psaArerverse: 5′-caggccacccttgaaggtg-3′) were designed using the primer design tool in SciTools on the Integrated DNA Technologies website; BLAST analyses (Altschul et al., 1990) against the Genbank nucleotide (nr/nt) database (http://www.ncbi.nlm.nih.gov/genbank/index.html) and Mushroom and Octopus Spring metagenomes (Klatt et al., 2011) were conducted to ensure that the top match of the primers was to *Synechococcus* spp. strain A or B′ *psaA* homologs.

Primers were diluted to a final concentration of 10 µM for use in the PCR reaction mixture (5 µl MgCl, 5 µl reaction buffer, 5 µl bovine serum albumin, 1 µl of each 10 µM primer stock solution, 1 µl dNTP's (10 µM each), 0.25 µl Taq Gold polymerase, and 31.75 µl ddH_2O). Cycling conditions were as follows: initial denaturing step of 94°C for 2 minutes followed by 30 cycles of 45 sec at 94°C, 45 sec at 53°C, and 90 sec at 72°C.

After verifying sizes of the PCR products using agarose gel electrophoresis, PCR products were purified using a QIAquick PCR purification kit (Qiagen, USA).

Purified PCR products were cloned using the TOPO TA 4.1 cloning kit (Invitrogen Life Technologies, USA) with 2 µl of DNA solution (~1 µg/ml), and cells were transformed following the manufacturer′s protocol. Transformant colonies were picked and inserts were analyzed by PCR amplification of the *psaA* locus and gel electrophoresis. Clones with correctly sized inserts were sequenced using the *psaA* forward and reverse PCR primers at the Idaho State University Molecular Biosciences Core Facility. Sequence data were aligned and manually edited using Sequencher 4.7. Clone sequence ends were trimmed to achieve a final length of 523 bp to obtain the maximum number of sequences for analysis. Sequences have been submitted to GenBank (JN160275 to JN160599). Neighbor-joining trees were constructed and average evolutionary divergence was calculated using MEGA 4.0 software (Tamura et al., 2007).

Ecotype Demarcation

Sequences were assigned as either *Synechococcus* A-like or B′-like at 60°C and 63°C if they were ≥92% identical to the respective *psaA* genomic homologs in *Synechococcus* spp. strain A (JA-3-3Aa) or B′ (JA-2-3B′a (2-13)). Aligned sequences for A-like and B′-like *psaA* variants were analyzed using ES to predict the number of putative ecotypes with 1.5x sorting (Cohan and Perry, 2007; Koeppel et al., 2008). Neighbor-joining trees were constructed and uploaded into ES as Newick files for ecotype demarcation. Sequences were classified into ecotypes using the automatic demarcation approach (http://fcohan.web.wesleyan.edu/ecosim/). ES demarcated putative

ecotypes as the largest clades that were consistent with containing a single ecotype (i.e., the confidence interval of the number of ecotypes for the clade included the value 1). AdaptML analysis was performed using software available at http://almlab.mit.edu/adaptml/ with temperature habitat variation, as well as the sequence-based phylogeny, as input (Hunt et al., 2008). Single sequences demarcated as PEs are identified with asterisks in Figures 2.1 and 2.2. Single sequences demarcated as ECs were not shown because the combination of nucleotide polymorphisms and habitat variation led to large numbers of singleton ECs that only contained a single nucleotide polymorphism. Singleton PEs and ECs are not considered in the main text so that distribution analyses could be focused on more abundant ecotypes that contain a dominant allele.

Denaturing Gradient Gel Electrophoresis

The DGGE procedure, band dissection and reagents were the same as in Ferris et al. (Ferris et al., 1996). DGGE gels were prepared to form a 50-60% denaturant gradient of urea and formamide, and gels were stained with SYBR Green and photographed with a Kodak 200R image station. DNA was then filtered through a QIAquick PCR purification kit (Qiagen, USA) to remove traces of acrylamide before PCR amplification and sequencing. Sequenced clones containing dominant alleles of closely related A-like PEs were PCR amplified and denatured to analyze co-migration patterns. DGGE patterns observed along the main flow path were replicated multiple times, and allelic distribution remained visibly unchanged from September 2006 to September 2008 (Appendix A Figure A3).

Physical/Chemical Analyses

The temperature of spring water at the sampling sites along the flow gradient was monitored using an electronic thermometer (DigiSense dual logger, Eutech Instruments, Singapore). Ambient photosynthetically active radiation (PAR, 400-700 nm) was quantified as downwelling irradiance as measured with a quantum irradiance sensor connected to a battery-driven data-logging light meter (LI 1400, LiCor, Nebraska, USA).

Profiles of O_2 concentration and pH versus depth in the mat were measured *in situ* with electrochemical microsensors mounted on a motor-driven micromanipulator and calibrated as described in detail elsewhere (Jensen et al., 2011, Steunou et al., 2008). Measuring sites were adjacent to and slightly upstream from mat collection sites so that debris from sampling would not affect the microsensor measurements.

Spectral photon scalar irradiance versus depth in the mat was profiled with fiber-optic microsensors as described in detail elsewhere (Kühl, 2005; Ward et al., 2006). Irradiance was measured in fresh mat samples retrieved at the different temperature sites with a core sampler (inner diameter ~1.5 cm) and brought to the laboratory for analysis within 24 hours after sampling. Light spectra were measured with a fiber-optic halogen lamp (KL2500, Schott GmbH, Germany) illuminating the mat vertically from above. A scalar irradiance microsensor was inserted into the mat at a zenith angle of 135° (relative to the vertically incident light) in steps of 100-200 µm vertical distance using a manually operated micromanipulator. The scalar irradiance microsensor was connected to a thermoelectrically cooled spectrometer (QE65000, Ocean Optics, USA).

Results

Variation in *psaA* and Prediction of Putative Ecotypes and Ecological Clusters

A highly variable region of the *psaA* gene was PCR amplified from DNA extracted from mat samples collected at Mushroom Spring 60°C, 63°C, and 65°C sites and the resultant amplicons were cloned. A sample of 245 clones from the 60°C site, 26 clones from the 63°C site, and 49 clones from the 65°C site were sequenced (184 B′-like and 136 A-like) to obtain the phylogenies shown in Figures 2.1 and 2.2. Compared to other loci studied previously, the average evolutionary distances for A-like and B′-like *psaA* variants were most similar to those of the *aroA* locus (Table 2.1). In separate analyses of *Synechococcus* A-like and B′-like populations, ES predicted a total of 21 non-singleton PEs (12 B′-like and 9 A-like), and AdaptML predicted a total of 14 non-singleton ECs (1 B′-like and 11 A-like). Most of the PEs and ECs contained dominant alleles, along with singleton variants, which were generally divergent from the dominant sequence by one or two nucleotide substitutions. ES estimated the periodic selection rate (sigma) to be 0.998 (confidence interval of 0.0195-2.13), and the ecotype formation event rate (omega) to be 0.152 (confidence interval of 0.104-0.342) (rates are per nucleotide substitution in the *psaA* sequence).

Of the twelve non-singleton B′-like PEs predicted by ES, nine contained a dominant variant (i.e., a sequence comprising >15% of a putative ecotype). PEs without a dominant variant always contained a small number of sequences (≤3). The B′-like

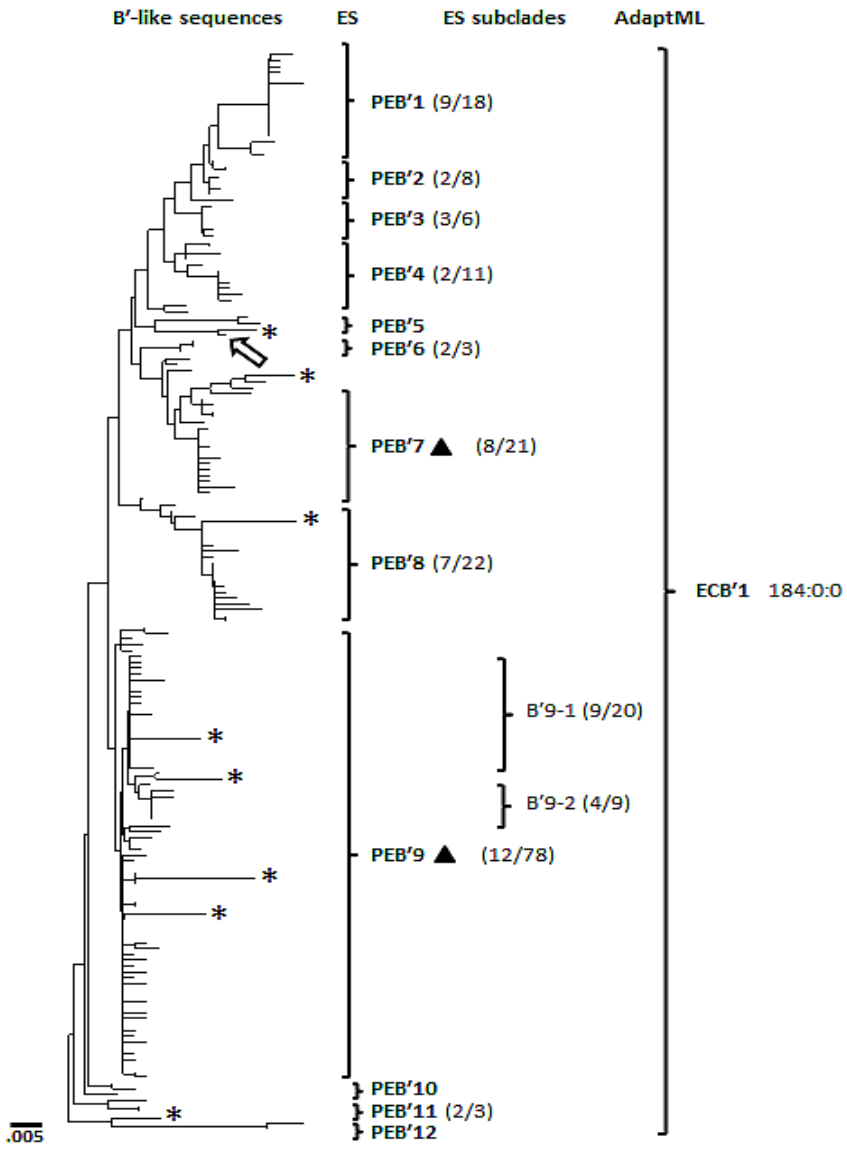


Figure 2.1. Neighbor-joining phylogenetic tree based on B′-like *Synechococcus psaA* sequences recovered from a 60°C site in the Mushroom Spring mat. Brackets show clusters of sequences demarcated as putative ecotypes (PEs) by Ecotype Simulation (ES) or, as an ecological cluster (EC) by AdaptML; subclades with dominant variants within PEs are also indicated. The ratio of the number of dominant sequence variants to the total number of variants in a PE or subclade is indicated within parentheses to the right of the PE designation. Triangle orientation indicates association with upper (▲) mat layers based on DGGE results. The location of the *Synechococcus* sp. strain B′ genomic homolog is shown by the arrow. Single sequences demarcated as PEs are indicated with asterisks. Scale bar indicates the number of fixed point mutations per nucleotide position.

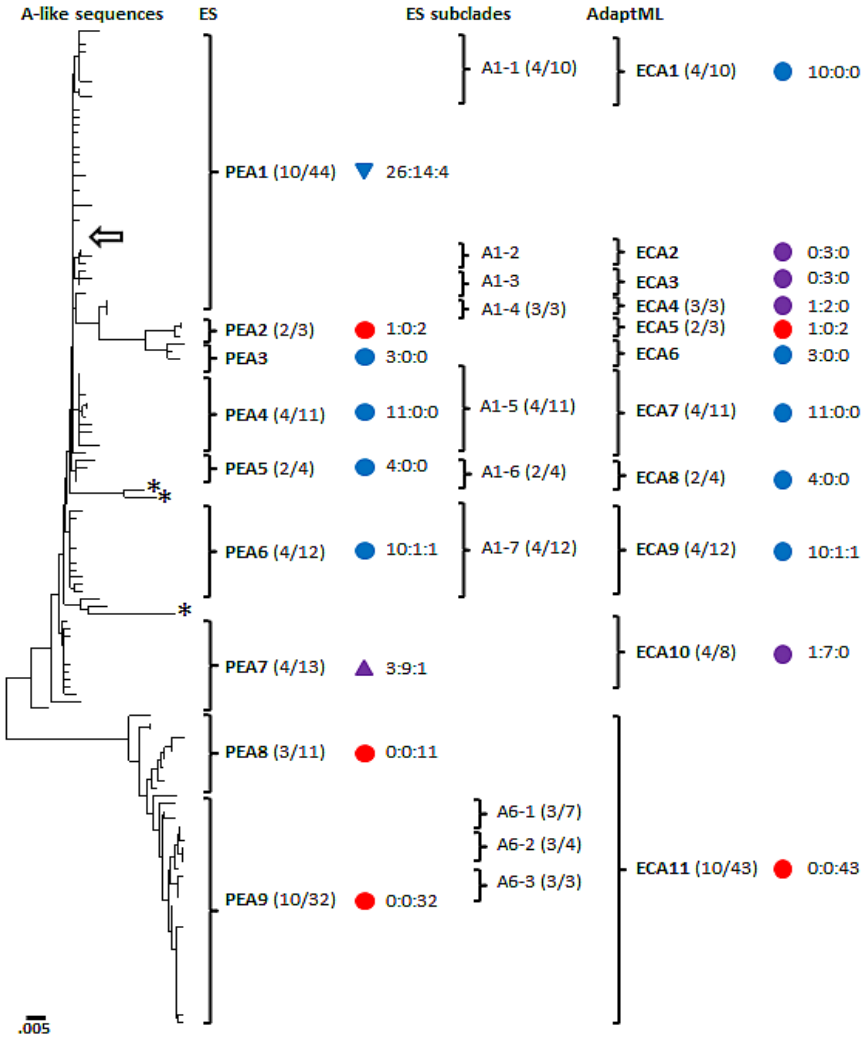


Figure 2.2. Neighbor-joining phylogenetic tree based on A-like *Synechococcus psaA* sequences recovered from 60°C, 63°C and 65°C sites in the Mushroom Spring mat. Brackets show clusters of sequences demarcated as putative ecotypes (PEs) by Ecotype Simulation (ES) or as ecological clusters (ECs) by AdaptML; subclades with dominant variants within PEs are also indicated. The ratio of the number of dominant sequence variants to the total number of variants in the PE, subclade or EC is indicated within parentheses to the right of the PE designation. Colored symbols indicate the temperature at which the majority of sequences within a putative ecotype were collected (blue, 60°C; purple, 63°C; red, 65°C); triangle orientation indicates association with upper (▲) and lower (▼) mat layers based on DGGE results. Numbers separated by colons to the right of some symbols indicate the number of sequences retrieved from each site (60°C: 63°C: 65°C). The location of the *Synechococcus* sp. strain A genomic homolog is shown by the arrow. Single sequences demarcated as PEs are indicated with asterisks. Scale bar indicates the number of fixed point mutations per nucleotide position.

Table 2.1. Molecular resolution and putative ecotype (PE) prediction by Ecotype Simulation (ES) analysis for *Synechococcus* A-lineage and B′-lineage populations from 60°C and 65°C Mushroom Spring samples based on the *psaA* locus and other loci using 70 sequences and allowing singleton PEs.

Locus[1]	AED[2] of A-like	AED of B′-	No. of A-like PE[3]s	No. of B′-like
16S	0.002	0.002	1	2
ITS	0.002	0.010	2	6
apcAB	0.016	0.010	5	5
aroA	0.013	0.030	4	13
psaA	**0.014**	**0.021**	**12(9)**	**18(12)**
rbsK	0.095	0.050	14	15

[1] loci other than *psaA* analyzed by Melendrez et. al. (23).

[2] AED, average evolutionary divergence.

[3] PE, putative ecotypes; PEs reported by Melendrez et al. 2011 were demarcated on the basis of variation in 70 sequences and included PEs demarcated based on singleton sequences. Comparable PE numbers for *psaA* variants were based on random samplings of 70 variants and included singleton PEs. Values in parentheses are numbers of non-singleton PEs predicted from all the *psaA* sequences included in this study, which are emphasized in the text.

sequences were also analyzed with AdaptML, which predicted one EC. Because B′-like sequences were recovered only from the 60°C sample; there was no opportunity for AdaptML to distinguish multiple EC's. The *Synechococcus* sp. strain B′ genome *psaA* sequence, which was not included in ES or AdaptML analyses, is most closely related to a singleton PE that fell between PE clades B′5 and B′6 (arrow in Figure 2.1). PE B′9 subclades (PE B′9-1 and PE B′9-2) did not contain enough molecular divergence for ES to demarcate them as unique PEs, though these groups of sequences each contain a dominant allele (Figure 2.1).

ES predicted nine non-singleton A-like PEs, eight of which contained a dominant allele. PEs A1 and A3-A6 were most frequently recovered from the 60°C sample, while PE A7 was mostly recovered from the 63°C sample, and PEs A2, A8 and A9 were almost

exclusively recovered from the 65°C sample. PE A8 and A9 are likely to be A′-like sequences, as they showed a closer relationship to homologs in a metagenome produced from DNA extracted from a 68°C mat sample (Appendix A, Figure A4), where A′-like 16S rRNA sequences predominate (Klatt et al., 2011). The *psaA* sequence from the *Synechococcus* sp. strain A genome was 100% identical to the dominant allele of the most abundant A-like PE (A1) (arrow in Figure 2.2). PEs A1 and A9 contain four and three subclades, respectively, each of which has its own dominant sequence variant, but these subclades were not demarcated as separate PEs by ES. The PE A1 subclades were all demarcated by AdaptML as unique ECs, one of which was only retrieved from the 60°C sample, and three of which were predominantly associated with the 63°C sample (EC A2 - EC A4). AdaptML did not resolve PE A8 and PE A9 and did not demarcate PE A9 subclades as unique ECs, but these groups were retrieved only from the 65°C sample (Figure 2.2).

Distribution of *psaA* Variants Along the Effluent Flow Path

Along the flow path between 65°C and 59°C, shifts in DGGE bands were observed at 64-65°C and 61-62°C (Figure 2.3). These shifts in population structure were observed at slightly higher temperatures in 2008 (Appendix A Figure A3), which might be due to differences in the degree to which the single measured temperature reflects the actual average temperature of a site (Appendix A Figure A2) or to year to year differences in water chemistry. Many of the DGGE bands were dissected, purified,

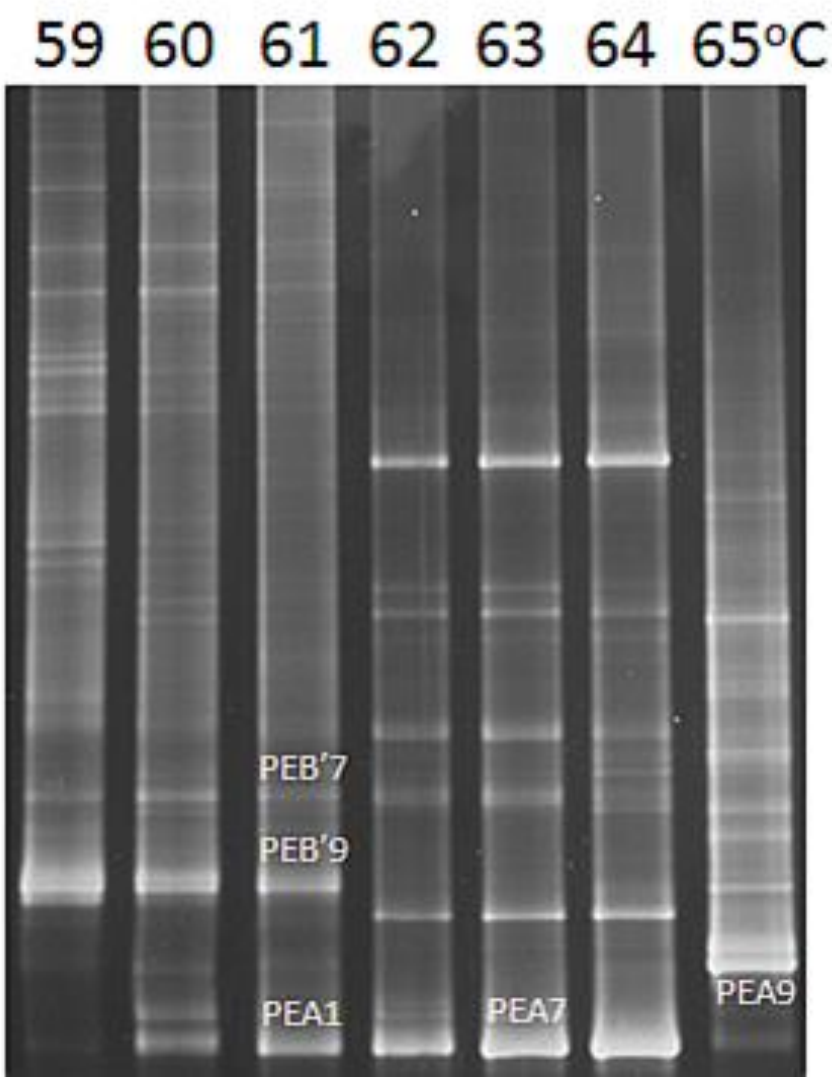


Figure 2.3. Denaturing gradient gel electrophoresis analysis of PCR-amplified *psaA* gene segments obtained from Mushroom Spring microbial mat samples collected at different temperature-defined sites (59°C to 65°C) along the flow path in the effluent channel. Labeled bands were purified, sequenced and found to correspond to the dominant variants of putative ecotypes B′7, B′9, A1, A7, and A9 shown in Figures 2.1 and 2.2. Direction of electrophoresis is top to bottom.

sequenced and analyzed with the clone library sequences to determine their relationship to PEs predicted by ES and ECs predicted by AdaptML. At 60-61°C, the three most intense DGGE bands corresponded to the dominant variants of the most well-represented

PEs in the B′-like and A-like phylogenies (B′9, B′7 and A1), which were defined by sequences of clones retrieved from the 60°C site. In the 63°C and 64°C samples, the band that migrated slightly lower in the gel than the A1 band corresponds to the dominant variant of PE A7 (EC A10), whose dominant allele was recovered only from the 63°C clone library. Bands corresponding to the dominant variants of PEs A1 and A7 exhibit how co-migration limits band resolution in DGGE analysis. Multiple distinct DGGE bands were observed at 65°C, the most intense of which corresponded to the dominant variant of the most abundant PE in the 65°C clone library (A9). A second dominant band migrated just above the dominant A9 band, and at least 3 other intense DGGE bands in the 65°C DGGE sample were visible that could not be purified and sequenced, but might correspond to PE A8 and subclades of PE A9, respectively. Alternatively, these bands may be heteroduplexes of these very closely related A and B′-like populations, since they also exhibit unique distributions along the effluent channel.

Vertical Distribution of *psaA* Variants

PCR-amplified *psaA* sequences obtained from different vertical sections of 60°C and 63°C mat samples were analyzed by DGGE to observe vertical population structure (Figure 2.4). Bands at different vertical intervals were identified based on co-migration with bands in the whole mat sample that had been purified and sequenced. At 60°C (Figure 2.4A), bands corresponding to the dominant variants of PEs B′9 and B′7 were most intense towards the top of the mat and decreased in intensity with depth, with the band corresponding to PE B′7 most intense closer to the mat surface and its intensity

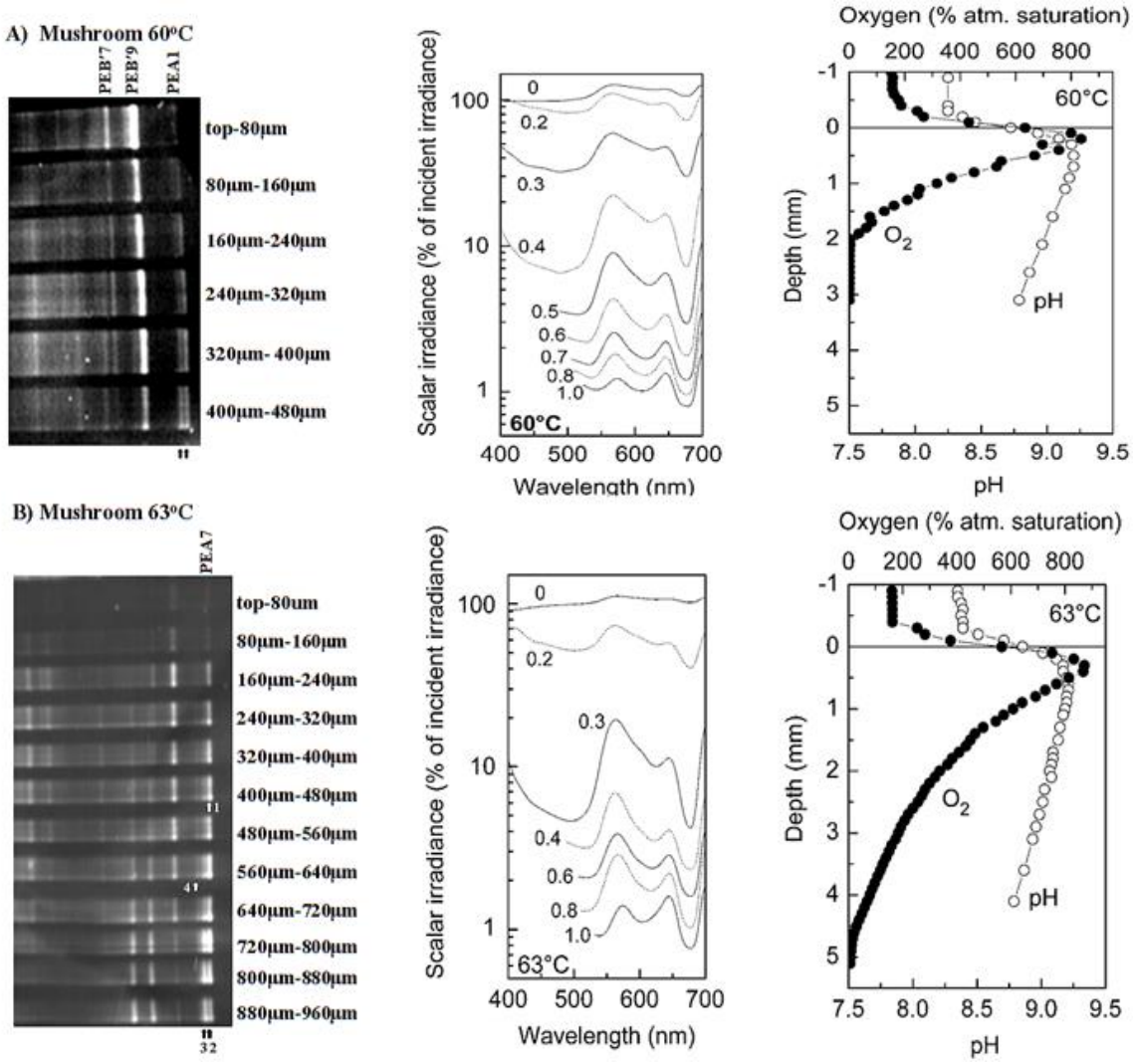


Figure 2.4. Vertical distribution of *psaA* variants, light, O_2 and pH at (A) 60°C and (B) 63°C sites in the microbial mat inhabiting the effluent channel of Mushroom Spring. (Left column) Denaturing gradient gel electrophoresis analyses of PCR-amplified *psaA* gene segments obtained from 80 µm-thick vertical sections of the top green layers. Arrows indicate the different positions of closely migrating dominant alleles which appear as one band in whole mat samples and co-migrate with the PE A7 band that was purified and sequenced (Figure 2.3). Direction of electrophoresis is left to right. (Middle column) Spectral scalar irradiance (400-700 nm). Data were normalized to the downwelling irradiance at the mat surface. The spectral coverage in deeper layers was narrowed due to spectrometer stray light at wavelengths <500 nm. (Right column) *In situ* depth profiles of O_2 concentration (solid symbols) and pH (open symbols) measured at noon (between 1200 and 1300 h at an irradiance of >1600 mmol photons m^{-2} s^{-1}).

decreasing more rapidly with depth than that of the band corresponding to PE B′9. Two distinct bands migrating to positions near that of the dominant variant of PE A1 were most intense in the lower layers of the 60°C sample (arrows in Figure 2.4A). It was not possible to separately purify and sequence these bands; however, we suspect that they are most likely from PEs A1 (and subclade A1-1, which is the same as ECA1), A4 and A6, which were the next most abundant PEs detected in cloning and sequencing analysis of the 60°C sample. Further support for this interpretation comes from the very similar DGGE migrations of the dominant alleles of these populations in DGGE analysis (Figure 2.5).

Several closely migrating bands were also observed in the 63°C vertical profile (Figure 2.4B). A band corresponding to PE A7 was found between the mat surface and a depth of 480 μm (band 1 in Figure 2.4B). Three other bands, which migrated slightly less far into the DGGE gel, were found at different subsurface depth intervals. The most intense and furthest migrating of these bands was found below a depth of 480 μm (band 2). Another band appeared between 400 μm and the bottom of the mat (band 3). The least intense of these bands, which exhibited the shortest migration distance, appeared between 240 and 720 μm (band 4). We suspect that bands 2, 3 and 4 represent PE A1 subclades A1-2, A1-3 and A1-4, as these were minor populations demarcated by AdaptML, which were demarcated as ECs A2, A3 and A4, and were predominantly retrieved from the 63°C site. Other dominant bands in the 63°C sample that could not be purified or sequenced may be heteroduplexes of these very closely related A-like populations, since

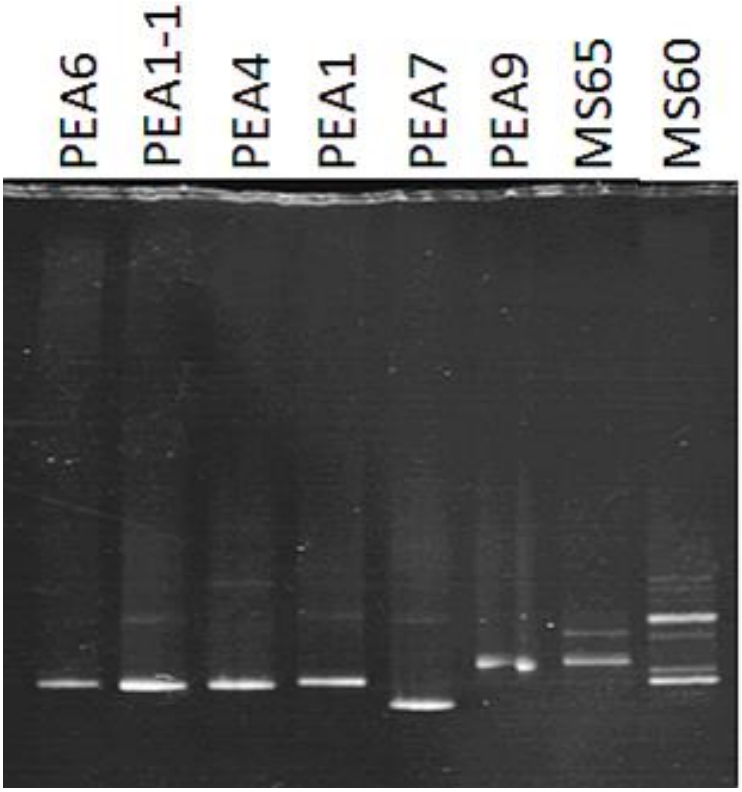


Figure 2.5. Denaturing gradient gel electrophoresis analyses of PCR-amplified *psaA* gene segments obtained from clones containing dominant alleles of closely related A-like putative ecotypes and mat DNA from 65°C and 60˚C (MS65 and MS60, respectively).

they also exhibit increasing band intensities with depth. The habitat distribution of each ecotype is summarized in Table 2.2 (see below).

Microenvironmental Characteristics

The microbial mats at the 60°C and 63°C sites both exhibited steep gradients of O_2, pH and irradiance with depth, leading to microenvironmental niche opportunity in the mats (Figure 2.4). During mid-day, when samples for molecular analyses were taken, intense solar irradiance fueled cyanobacterial photosynthesis in the upper mat layers, where high pH (>9) and strong O_2 supersaturation was observed in the top 1 mm (Figure

2.4 right panel). Maximal O_2 concentration, indicative of peaks in gross photosynthesis, were found at 0.2 mm and 0.4 mm below the mat surface at the 60°C and 63°C site, respectively. At 60°C, O_2 was fully depleted 2 mm below the mat surface, while O_2 penetrated to about 5 mm in the 63°C mat.

Spectral light measurements showed distinct troughs in the transmission spectra corresponding to absorption maxima of Chl *a* (around 440 nm and 675 nm) and cyanobacterial phycobiliproteins (especially phycocyanin absorbing around 625 nm) in the upper millimeters of the mat (Figure 2.4 central panels). Over 0.5 to 1 mm below the surface of the mat the phycobiliprotein absorption maximum gradually changed to about 600 nm. Light in the visible spectral region driving oxygenic photosynthesis (400-700 nm) was depleted to about 1% of the incident irradiance 1 mm below the mat surface in both mats. Light attenuation was highest in the top 0.3-0.5 mm where a peak in O_2 concentration and pH was observed. A very strong attenuation was observed between 0.2 and 0.3 mm below the mat surface at the 63°C site, indicative of a densely pigmented subsurface cyanobacterial layer. At the 60°C site, light attenuation was somewhat more gradual and was maximal between 0.3 and 0.4 mm below the surface.

Discussion

Identification of the fundamental species-like units within a microbial community is of paramount importance to understanding how evolutionary forces have shaped bacterial lineages, defining their adaptations and thereby controlling their present distributions and contributions to the community as a whole. Although many models of

species and speciation have been developed (Cohan and Perry, 2007), the Stable Ecotype Model seems to best fit these *Synechococcus* spp. populations (Ward, 1998; Ward et al., 2006; Ward and Cohan, 2005).

Melendrez et al. (2011) showed that the number of PEs detected by ES depends on the molecular resolution of the protein-encoding locus used to sample genetic diversity and the fraction of diversity sampled. In order to compare results from *psaA* analyses with those of Melendrez et al. (2011), we had to reduce the number of sequences analyzed to 70 and include singleton PEs in the analysis (Table 2.1). In terms of the average evolutionary divergence among the sequence variants studied, *psaA* offers more molecular resolution than do 16S rRNA, ITS and *apcAB* loci, similar molecular resolution as the *aroA* locus, but less molecular resolution than the *rbsK* locus (Table 2.1). The total number of A-like and B′-like PEs predicted by ES using these common conditions (i.e., 70 sequences and allowing singleton PEs) increased from 3 using 16S rRNA, to 8 using the ITS locus, to 10 using the *apcAB* gene, to 17 using the *aroA* gene, to 30 using the *psaA* locus. ES analysis of the *rbsK* locus, which exhibited significantly more evolutionary divergence, identified 29 ecotypes in the *Synechococcus* A/B lineage found at 60°C and 65°C. The similar number of PEs detected using *psaA* and *rbsK* variation might be taken to indicate that both loci offer sufficient molecular resolution to detect all PEs. However, it is difficult to draw this inference for two reasons. First, our analysis of PEs using *psaA* involved three temperature sites, whereas only two sites were included in the analysis of other loci, and some PE's found at the third temperature (i.e., 63˚C) were unique to that site. Second, PEs likely to be contributed by A′-like

Synechococcus dominated the 65°C site we studied, whereas they comprised a small proportion of sequences in the 65°C site studied by Melendrez et al. (2011), which were collected in 2004. Suspected A′-like sequences were not included in analyses done by Melendrez et al. (2011). Differences in population structure at 65°C in 2006 and 2008 (Figure 2.3 and Appendix A, Figure A3) may have been due to the fact that single temperature readings are not indicative of the true range of temperatures at a given sample site (Appendix A, Figure A2). It is also unclear without deeper sampling whether singleton PEs represent actual ecotypes. Consistent with previous studies (Melendrez et al., 2001), the estimates of 0.998 for periodic selection rate (sigma) and of 0.152 for ecotype formation events (omega), indicated a higher rate of periodic selection than ecotype formation, as predicted by the Stable Ecotype Model.

ES analysis, which is based solely on sequence variation, predicted PEs in cases where AdaptML did not, due to low habitat variation in the B′-like phylogeny (Figure 2.1). For instance, the twelve B′ PEs predicted by ES were all lumped into one EC by AdaptML, because B′-like sequences were found only in the 60°C sample. Also, the inability of AdaptML to discern PEs A8 and A9 and subclades of PE A9 reflects the fact that all sequences in these PEs were obtained only from the 65°C sample. In contrast, given sufficient habitat variation, AdaptML predicted ECs that exhibited too little sequence variation for ES to demarcate as PEs. This is exhibited by the fact that AdaptML predicted four ECs within PE A1, discerning the closely related subclades not demarcated by ES analysis. This supports our hypothesis that the true ecotype populations are the smallest clades containing a single dominant variant. Clearly, both

molecular resolution and habitat variation are important factors for identifying ecological species.

The abrupt shifts of dominant variants of several PEs at the same locations (e.g. at 64-65°C and 61-62°C in 2006) along the flow path may indicate that common molecular adaptations underlie temperature ranges of co-existing species-like populations. These transition zones may relate to ecotones (i.e. a transitional zone between two ecological communities, as between a forest and grassland or a river and its estuary), except on a microbial scale. These transition zones are apparent along effluent flow and vertical profiles, indicating that even in such a small, relatively simple environment multiple ecotones are present and influencing the evolution and ecology of the mat community as a whole.

DGGE distribution analyses along flow and vertical gradients show patterns of variation at a level of molecular resolution higher than previously attained, but matching variation seen in DGGE gels with ecotypes predicted from the cloned sequences was limited to abundant ecotypes that contain a dominant allele that did not co-migrate with other dominant alleles. Ecotypes identified by ES or ECs identified by AdaptML that contain a dominant allele appear to occupy different spatial niches, though many ecotype distributions have overlapping regions where they spatially co-exist (Table 2.2). Overlapping spatial niches were also reported in the studies of *Bacillus* and marine *Vibrio* ecological species (Conner et al., 2010; Hunt et al., 2008; Koeppel et al., 2008). At lower temperatures the B′-like ecotypes are most abundant and were detected primarily at the

Table 2.2. Summary of A-like, A′-like and B′-like *Synechococcus* putative ecotypes and subclades containing a dominant allele indentified by Ecotype Simulation, and ecological clusters identified by AdaptML, and their distributions along flow and vertical gradients.

ES[1] **PE**[2]**s**	**Subclade**	**DA ratio**[3]	**No. Sequences From** 60°C	63°C	65°C	**Vertical habitat**[4]	**AdaptML**[5]	**DA ratio**	**No. Sequences From** 60°C	63°C	65°C
PEA1		10/44	26	14	4	160-960µm					
	A1-1	4/10	10	0	0		**ECA1**	4/10	10	0	0
	A1-2	0/3	0	3	0		**ECA2**	2/3	0	3	0
	A1-3	2/3	0	3	0		**ECA3**	2/3	0	3	0
	A1-4	3/3	1	2	0		**ECA4**	3/3	1	2	0
PEA2		2/3	1	0	2		**ECA5**	2/3	1	0	2
PEA3		0/3	3	0	0		**ECA6**	0/3	3	0	0
PEA4		4/11	11	0	0		**ECA7**	4/11	11	0	0
PEA5		2/4	4	0	0		**ECA8**	2/4	4	0	0
PEA6		4/12	10	1	1		**ECA9**	4/12	10	1	1
PEA7		4/13	3	9	1	0-480µm	**ECA10**	4/8	1	7	0
PEA8		3/11	0	0	11		**ECA11**	10/43	0	0	32
PEA9		10/32	0	0	32						
PEB′1		9/18	18	0	0						
PEB′2		2/8	8	0	0						
PEB′3		3/6	6	0	0						
PEB′4		2/11	11	0	0						
PEB′5		0/3	3	0	0						
PEB′6		2/3	3	0	0						
PEB′7		8/21	21	0	0	0-80µm	**ECB′1**	-	184	0	0
PEB′8		7/22	22	0	0						
PEB′9		12/78	78	0	0	0-400µm					
PEB′10		0/3	3	0	0						
PEB′11		2/3	3	0	0						
PEB′12		0/2	2	0	0						
Singletons		0/11	11	0	0		**Singletons**	0/10	7	1	2

[1] ES, Ecotype Simulation.

[2] PE, putative ecotypes. Dashed numbers indicate subclades within a demarcated PE that contain a dominant allele, but were not demarcated as a putative ecotype.

[3] ratio of dominant allele (DA) number to total number of sequences within a putative ecotype.

[4] Vertical depth intervals are for PEs indicated only. PEs without vertical depth intervals could not be indentified in gel analysis.

[5] EC, ecological clusters.

top of the mat, which was consistent with previous research conducted with lower resolution loci (Ramsing et al., 2000). However, the ability to discern B′-like ecotype distributions was limited since they were recovered from only one temperature site. The A-like ecotypes were recovered from all three temperature sites, and clusters of sequences identified as ecotypes by ES, subclades with dominant variants, and ECs identified by AdaptML showed different distributions. The vertical analysis of 60°C and 63°C samples showed multiple bands in the lower layers, which we hypothesize are associated with closely related PEs or ECs (i.e., PE A1 subclades) containing dominant alleles that could not be differentiated within the whole mat sample by DGGE due to co-migration. We speculate that these deeper populations could possibly align with the observation of the phycobiliprotein absorption maximum shift to about 600 nm in deeper mat layers. Ecological populations of both A-like and B′-like *Synechococcus* cyanobacteria are specialized to very fine gradations of temperature (horizontal flow position) and parameters that vary with depth (e.g. light intensity, light quality and/or chemical gradients).

In order to achieve increased sampling of diversity over a larger variety of habitats, we are currently examining the use of Ti454 bar-code analyses (Andersson et al., 2008; Mehrdad et al., 2007), which will also alleviate co-migration problems by providing sequence data that discriminate variants based on single nucleotide polymorphism patterns. This will permit clearer distinction of dominant variants of PE clades, based on much larger numbers of variants observed in a much larger number of habitats, and allow us to determine if the single sequences predicted to be ecotypes represent additional rare ecologically distinct populations. We are also using bar-code

analyses to examine the possibility of co-existing PEs that are differently adapted to temporally varying environmental parameters. Such PEs might, for instance, be adapted to nutrients like Fe^{+2} or fixed N forms whose availability is discontinuous throughout a diel cycle (Bhaya et al., 2007; Cohan and Perry, 2007; Klatt et al., 2011). Since spatial co-occurrence does not test whether the members of PEs are ecologically interchangeable, we are also currently conducting population dynamic studies to observe whether or not PE populations change as an ecological unit in response to environmental perturbations. Finally, we are investigating the alternative explanation that horizontal gene exchange may have caused >1 dominant alleles to occur within individual ecological populations (Melendrez et al., in prep).

Acknowledgements

This research was supported by the National Science Foundation Frontiers in Integrative Biology Research Program (EF-0328698), the National Aeronautics and Space Administration Exobiology Program (NAG5-8807), the Danish Natural Science Research Council (MK, SIJ) and the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program 395 (GSP) [This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL); contract #112443]). This study was conducted under Yellowstone National Park research permits YELL-0129 and 5494 (DMW) and YELL-02058 (MK) and we appreciate the assistance from National Park Service personnel. We would also like to thank for Alex Koeppel for his assistance in answering questions and troubleshooting protocols for Ecotype

Simulation, Jason Wood for his optimization of Ecotype Simulation, Lawrence David for development of the AdaptML program and Jane Wiedenbeck for assistance with the interpretation of AdaptML results.

CHAPTER THREE

PYROSEQUENCING ANALYSES OF *SYNECHOCOCCUS* ECOLOGICAL SPECIES INHABITING THE MICROBIAL MAT OF MUSHROOM SPRING, YELLOWSTONE NATIONAL PARK

Contributions of Authors

Manuscript in Chapter 3

Author: Eric D. Becraft

Contributions: Designed the study, collected samples, conducted molecular analyses, analyzed output data, and primary writer of the manuscript.

Co-Author: Fred M. Cohan

Contributions: Mentor and developer of evolutionary simulation program used to demarcate ecological species.

Co-Author: Michael Kühl

Contributions: Microsensor analysis and chemical measurements (e.g. pH and light).

Co-Author: Sheila I. Jensen

Contributions: Microsensor analysis and chemical measurements (e.g. pH and light).

Co-Author: Jason M. Wood

Contributions: Program designer

Co-Author: Doug B. Rusch

Contributions: Collaborator at the J. Craig Venter Institute. Primary overseer of high-throughput sequencing.

Co-Author: David M. Ward

Contributions: Primary mentor, laboratory provider and co-writer.

Manuscript Information Page

Eric D. Becraft, Frederick M. Cohan, Jason M. Wood, Doug B. Rusch, Michael Kühl, Sheila I. Jensen and David M. Ward

International Society for Microbial Ecology Journal

____ Prepared for submission to a peer-reviewed journal
__x_ Officially submitted to a peer-review journal
____ Accepted by a peer-reviewed journal
____ Published in a peer-reviewed journal

Abstract

Previous analyses of sequence variation in the protein-encoding gene *psaA* showed that putative ecological species (putative ecotypes) predicted by Ecotype Simulation had unique distributions along flow and vertical gradients in alkaline siliceous hot spring microbial mats. However, the previous analyses were limited in their ability to deeply sample and resolve sequence variants in a large variety of habitats. In this study Ti454-barcode high-throughput sequencing was used to 1) sample more habitat types, 2) obtain deeper sequence coverage per habitat, and 3) sample greater sequence variation within predicted ecological species populations. Ecotype Simulation was used to demarcate putative ecotypes across all high-frequency sequences and microbial mat habitats. More putative ecotypes were detected than in previous studies, and most exhibited unique distributions along the effluent channel and with depth in the mat. Putative ecotypes with different vertical distributions differed in temporal patterns of expression of the *psaA* gene, consistent with patterns of light availability at each depth. Putative ecotypes responded uniquely to experimental perturbations of temperature and light environment. Genetic variation predicted to be contained within putative ecotypes was maintained as the relative abundances of ecotypes changed, indicating that each population responded as a set of ecologically interchangeable individuals. Thus, Ti454-barcode analyses demonstrated that putative ecotypes predicted by Ecotype Simulation have the properties expected of ecological species.

Introduction

Studies of diverse microbial habitats have shown the existence of multiple closely related populations that are distributed differently along environmental gradients (West and Scanlan, 1999; Béjà et al, 2001; Ward and Cohan, 2005; Cohan et al, 2006; Johnson et al, 2006; Ward et al, 2006; Walk et al, 2007; Hunt et al, 2008; Manning et al, 2008; Martiny et al, 2009; Connor et al, 2010; Denef et al, 2010; Shapiro et al, 2012). These observations have prompted the realization that some microorganisms might exist as ecological species (Ward, 1998; Ward and Cohan, 2005), because they meet one of the basic expectations that ecological species occupy distinct niches. However, for several reasons, ecological distinction alone may not provide sufficient evidence of ecological species populations. First, while molecular resolution may be sufficient to discern ancient adaptations, it may be too low to detect younger, more closely related descendant species whose subsequent adaptations require greater molecular resolution to detect (Ferris et al., 2003; Martiny et al, 2009). This is especially problematic with slowly evolving genetic markers like 16S rRNA. Second, ecological species with different behavioral ecologies might share a single spatial niche (e.g., nocturnal and diurnal eukaryotic species). Finally, theory predicts that the members of an ecological species population are ecologically interchangeable (Cohan, 2011; Koeppel et al, 2013). In other words, each member of an ecological species population would have the same opportunity to perpetuate the species during its evolution.

Earlier molecular studies of hot spring microbial mat communities provide an example of the problem of limited molecular resolution. 16S rRNA analyses revealed cyanobacterial variants A″, A′, A, B′ and B *Synechococcus*, which exhibited different

distribution patterns between 72-74°C and ~50°C along the effluent channel flow path (Ferris and Ward, 1997), or in vertical position within the photic zone of the mat (Ramsing et al, 2000). However, subsequent studies revealed the need for higher molecular resolution to detect ecological species populations (Ferris et al, 2003). Population genetics analyses (Melendrez et al, 2011), and fine-scale distribution analyses of *Synechococcus* sequence diversity along effluent flow and vertical gradients (Becraft et al, 2011) using protein-encoding loci showed how molecular resolution influences identification of ecological species populations in the Mushroom Spring mat. In these studies, rather than using an arbitrary molecular divergence cutoff, an evolutionary simulation algorithm based on the Stable Ecotype Model of species and speciation was used to predict ecological species. The algorithm, Ecotype Simulation (Koeppel et al, 2008), hypothesizes, from sequence variation alone, the phylogenetic boundaries of ecological populations, each representing a set of ecologically interchangeable individuals. Each of these populations is considered a putative ecotype (PE), because the output of the algorithm is hypothesized ecological species, whose ecological distinctions and ecologically interchangeable membership have not been demonstrated. Ecotype Simulation models the sequence diversity within a bacterial clade as the evolutionary result of net ecotype formation, periodic selection (Koch 1973) and drift, yielding the number of PEs within an environmental sample; Ecotype Simulation also identifies the sequence variants grouped into each PE.

In Becraft et al. (2011), genetic variation at the *psaA* locus was analyzed because (i) it offers more molecular resolution than the 16S rRNA locus, (ii) it is known to be expressed *in situ* (Liu et al, 2012), (iii) it exists as a single copy in the genomes of mat

Synechococcus isolates (Bhaya et al, 2007), (iv) the protein subunit it encodes (PsaA) is necessary for photosynthesis, and (v) has shown no evidence of recombination in the region studied. Sequences of *psaA* retrieved by PCR amplification, cloning and sequencing from samples collected along the effluent flow channel were demarcated into *Synechococcus* PEs. Since PEs contained dominant variant sequences we used denaturant gradient gel electrophoresis (DGGE) to demonstrate their distinct distributions along flow and vertical gradients. However, co-migration of dominant variants in DGGE analyses made band purification and sequencing difficult, and both sampling of variation and the number of habitat samples analyzed were limited. Furthermore, because these studies did not consider all variants within a putative ecotype, they did not address the problem of ecological interchangeability.

In this study, we used Ti454-barcode pyrosequencing to overcome these limitations resulting in a more complete, sequence-based view of the genetic and ecological diversity within the *Synechococcus* populations of this community. Pyrosequencing has been shown to overestimate the diversity within microbial populations because of a high rate of sequence errors (Reeder and Knight, 2009). However, since reference clone library sequences and genomes of representative *Synechococcus* isolates were available (Allewalt et al, 2006; Bhaya et al, 2007; Becraft et al, 2011), it was possible to recognize such artifacts by alignment to known reading frames. To test the ecological distinctions of PEs we (i) extended fine-scale distribution studies, (ii) examined PE-specific transcription patterns over a diel cycle, and (iii) investigated the responses of *Synechococcus* populations to environmental perturbations. Additionally, since the greater depth of coverage led to recovery of multiple sequence

variants within PEs, responses to perturbations allowed us to test for ecological homogeneity within each PE, based on the expectation that all members of an ecological species population would change uniformly in response to environmental perturbations. We provide evidence that many of the most abundant PEs predicted by Ecotype Simulation are ecologically distinct from one another and that the individuals of a given PE are ecologically homogeneous.

Materials and Methods

Sampling

This study was focused on the ~60 to 68°C region of the major effluent channel of Mushroom Spring, an alkaline siliceous hot spring in the Lower Geyser Basin, Yellowstone National Park, WY, USA. Samples were collected from sites spread over a distance of ~10 m along the main flow path defined by temperatures measured at the time of collection (see Appendix A). Mat samples for distribution studies were collected on 12, 13, 14 and 15 September 2008 (60, 63 and 65°C), and 12 and 13 September 2009 (68°C) using a #2 cork borer (19.6 mm^2) and were immediately frozen in liquid N_2. Cores for vertical analysis were instantly frozen in isopentane cooled with liquid N_2 to minimize decomposition of nucleic acids and to preserve core integrity. These cores were subsequently dissected at ~80 µm intervals using a cryotome, as described in Becraft et al. (2011). Samples for transcript analysis over a full diel cycle were taken at hourly intervals starting at 0500 h on 11 September 2009, and continuing until 0400 h on 12 September 2009 using a #4 cork borer (38.5 mm^2). Replicate samples were collected over a ~1 m^2 area then pooled.

Perturbation Experiments

Temperature Shift. Samples for temperature-shift experiments were collected in replicate using #2 cork borers, placed into 3 ml glass perfume vials that were filled with spring water flowing above this region of the mat, capped, and suspended in the effluent channel at a higher temperature site by aluminum wires attached to wooden stakes on either side of the channel (see Appendix B Figure B1). Duplicate samples were retrieved two and four days after the disturbance was initiated on 27 October 2008. Previous studies showed that confining samples in this way did not alter the initial population structure of samples when incubated at the collection site (i.e., not shifted in temperature) (Ruff-Roberts et al, 1994). Some bleaching was noticed in later stages of the incubation period, likely due to the altered flow in the closed vials which severely impacts mass diffusion, or other inhibiting factors.

Light Shifts. On 27 October 2008, two 254 cm^2 rectangular wooden platforms were placed ~2.5 cm above the mat surface at a ~63°C site. One was covered with 4 layers of stretched muslin to reduce ambient light by ~92%. The other was covered with UV-blocking Plexiglas (UF-5; 3.2-mm thick; Plexiglass: Autohaas N. Am., Philadelphia, PA, USA), which transmits ~90% of the visible spectrum, but only ~5% of the near-UVA spectral region and no measurable light in the UVB region (Norris et al, 2002) (Appendix B Figure B1). Samples were immediately frozen on dry ice (-20°C) in the field and kept frozen at -80°C until analysis. Unaltered controls are described in Appendix Section B IV.

Molecular Methods

DNA was extracted and purified as described in Becraft et al. (2011), and RNA was extracted and purified as described in Liu et al. (2011). Primers for the amplification of *Synechococcus* A/B-lineage *psaA* genes (psaAcenterforward: 5′-TTCCACTACCACAAGCGGGCTCC-3′, psaAreverse: 5′-CAGGCCACCCTTGAAGGTG-3′) were designed to yield a 324 bp segment to maximize the number of single-nucleotide positions that could be used to differentiate PEs, and were tested for *Synechococcus* A/B′ specificity as described in Becraft et al. (2011). Reduced sequence length caused subclades A1-3 and A1-4 identified in Becraft et al. (2011) based on 523 nt sequences to collapse into PE A1, and subclade A′9-2 to collapse into PE A′9 (Table 3.1), possibly indicating that PEs A1 and A′9 are inclusive of >1 ecotype.

To estimate ecotype-specific expression patterns, *psaA* cDNA was synthesized from RNA (after DNase treatment) using psaAcenterforward primer (see above) with the SuperScript III First-Strand Synthesis Supermix (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions, and the product was PCR-amplified according to protocols described in Becraft et al. (2011). Controls with only RNA (i.e., without cDNA) were used during amplification to insure that all DNA had been degraded. Ti454-barcoding and sequencing (Chapter 1) were completed at the J. Craig Venter Institute according to the GS FLX Titanium Series Rapid Library Preparation Method Manual. DNA was sheared using the Covaris S2 System, and qPCR was used to accurately estimate the number of molecules needed for emulsion PCR

Identification of Sequences
Containing Homopolymeric Artifacts

The Ti454-barcoding sequence data were first processed to remove homopolymeric errors (base pair insertions or gaps following strings of identical nucleotides) and sequences of poor quality. Using ClustalW to generate alignments, an error-fixing Perl script (Homopolymer Extinguisher; custom scripts are available upon request) aligned each raw sequence (plus its complement, reverse, and reverse complement sequences) with an in-frame consensus reference sequence of all *psaA* sequences from Becraft et al., (2011) to determine the best possible alignment. Nucleotides in the raw sequences that caused a gap to form in the reference sequence reading frame with the best alignment of these sequences were assumed to result from erroneous homopolymers (Quince et al, 2009, 2011; Reeder and Knight 2009), and these nucleotides were excised. This step also trimmed the raw sequences to the same length as the reference sequence. Any trimmed raw sequences containing gaps relative to the consensus reference sequence were removed from further analysis. This yielded 164 467 sequences, with an average of 1 713 (standard deviation of 251) for each of 96 unique environmental samples. Separately, 17 total cDNA samples yielded, on average, 360 sequences per sample (standard deviation of 120). cDNA sequences were trimmed to a 247 bp segment to obtain the maximum number of sequences for analysis.

Ecotype Demarcation

It was too computationally challenging to analyze all sequence variation using the current version of Ecotype Simulation. Thus, a custom Perl script algorithm (High Frequency Sequence Finder) was used to generate a list of high-frequency sequences by

counting the number of occurrences of each unique sequence. High-frequency sequences were defined as sequence types numbering more than 50 identical representatives across all samples (96 DNA and 17 cDNA), which included all PE dominant variants identified in Becraft et al. (2011). High-frequency sequences showed no evidence for recombination using the RDP3 software package (Martin et al., 2010). Each high-frequency sequence was assigned as either *Synechococcus* A-like or B′-like if it was ≥95% identical to the respective *psaA* genomic homologs in *Synechococcus* spp. strain A (JA-3-3Aa) or B′ (JA-2-3B′a (2-13)) (Bhaya et al, 2007). Sequences suspected to correspond to the A′-like 16S rRNA sequence (Ferris and Ward 1997), and showing the highest similarity to 68°C metagenomic library sequences, which contains predominantly A'-like variants (see Appendix B Figure B4; Klatt et al, 2011), were included in the A-like phylogeny. Single copies of each unique high-frequency sequence were pooled (separately for B′-like and A plus A′-like lineages) and high-frequency sequences of each lineage were analyzed separately using Ecotype Simulation (with 1.5x sorting) to predict the number of PEs (Cohan and Perry, 2007; Koeppel et al, 2008). Neighbor-joining trees were constructed and uploaded into Ecotype Simulation as Newick files for ecotype demarcation. Two methods for Ecotype Simulation ecotype demarcation have been developed (Melendrez et al, 2011; Becraft et al, 2011; Francisco et al, 2012). The more conservative approach tends to yield ecotypes that are more inclusive clades; here an ecotype is demarcated as the most inclusive phylogenetic group for which the confidence interval for the number of predicted ecotypes includes the value 1. In our alternative, fine-scale demarcation, ecotypes are demarcated as the largest groups for which the maximum likelihood solution for the number of predicted PEs is the value 1. B'-like high-

frequency sequences were phylogenetically more divergent and were demarcated using the conservative demarcation approach, while the less divergent A-like and A'-like high-frequency sequences were demarcated using the fine-scale approach (see Appendix B Section I and Appendix Figure B2 for comparisons of fine-scale and conservative demarcations). The Ecotype Simulation software and instructions for its use are freely available at http://fcohan.web.wesleyan.edu/ecosim/.

Determination of All Variants within Putative Ecotype Populations

To expand the number of sequences associated with each high-frequency sequence, a neighbor-joining phylogenetic tree was created for all unique A-like plus A′-like, or B′-like sequences using Phylip′s dnadist and neighbor-joining programs. Any low-frequency sequence (<50 copies in all samples) falling into the same clade as a single high-frequency sequence was then grouped with that sequence. Finally, a matrix of the number of occurrences of each high-frequency sequence (and expanded clades) in each sample was generated for further analysis. The number of dominant variant sequences, other high-frequency sequences predicted to belong to a PE, and the low-frequency sequences associated with them were summed to obtain the total number of variants detected for each PE in each sample (PE sequences = dominant variant sequence + all other high-frequency sequences + all low-frequency sequences). The percentage contribution of each PE in each environmental sample was calculated by dividing the total number of sequences within a PE detected in a sample by the total number of *psaA* sequences in that sample ((PE sequences in sample / all associated sequences in sample)

x100). Absolute quantitation of PE abundance was not achieved in these analyses because PCR and sequencing only samples proportions of populations in the data.

Statistical Analyses

G-tests were conducted on *psaA* sequence counts along all environmental gradients for abundant PEs, and for A-like and B′-like PEs separately, to determine whether there was statistical significance for unique PE distributions (see Appendix Table B1). Ancova tests were conducted for replicated perturbation experiments to determine whether PE populations differed in their response to environmental change (see Appendix Table B2). G-tests and Ancova analyses were done in Stata (StataCorp LP, Stata Statistical Software: Release 13, College Station, TX). Additionally, over the course of the light reduction and temperature shift experiments we tested whether PEs changed their relative frequencies, and we tested whether the membership of high-frequency sequences (and associated low-frequency sequences) in a given PE changed in unison. This was done by evaluating whether the percent of each high-frequency sequence within a PE, and the ratio of low-frequency sequences to high-frequency sequences within a PE, changed separately over time using a generalized linear model (a flexible generalization of ordinary linear regression that allows for binomial distributions; see Appendix B Table B3).

Microsensor Measurements

Diel sampling for molecular analyses was closely coordinated with simultaneous logging of the incident downwelling solar irradiance and *in-situ* microsensor measurements of O_2 concentration profiles in the microbial mats at the sample site

following the same calibration and measurement procedures described in detail in earlier studies (Becraft et al. 2011, Jensen et al. 2011). At the end of the diel measurements, ~3 cm^2 of mat was sampled with a glass corer and transported to the laboratory in spring water for subsequent measurements (within 24 hours) of spectral light penetration using fiber-optic scalar irradiance microsensors (see details in Kühl 2005, Ward et al. 2006).

Scalar irradiance spectra measured in particular depths of the mat were normalized to the incident downwelling irradiance at the mat surface and expressed in % of downwelling irradiance at the mat surface. Additionally, measured scalar irradiance spectra in each depth were integrated over 400-700 nm (i.e. photosynthetic available radiation (PAR)) and normalized to the downwelling irradiance of PAR at the mat surface. This depth profile of E_0(PAR) (in % of downwelling irradiance) was then used together with the actual measured downwelling photon irradiance at the field site during the diel study to construct depth profiles of *in situ* E_0(PAR) in units of µmol photons m^{-2} s^{-1} for all sampled time points in the diel study.

Isopleth diagrams of O_2 concentration and photon scalar irradiance, E_0(PAR) depth distribution over the diel cycle were constructed from depth profiles measured at different time points using the graphing software Origin Pro 8.5 (Origin Lab Corp., Northampton, USA).

Results and Discussion

Ecotype Simulation Analysis

A total of 118 high-frequency sequences were found in the variation detected in 113 samples (96 DNA; 17 cDNA) and were analyzed using Ecotype Simulation to

predict PEs (Figure 3.1). Most PEs contained multiple high-frequency sequences (Figure 3.1), and each high-frequency sequence also contained multi-copy low-frequency sequence variants as well as singleton sequences (not shown in Figure 3.1). Low-frequency sequences were the result of one or two base-pair differences from high-frequency sequences. Dominant variants which represented the plurality of sequences in a PE were closely related to high-frequency sequences containing fewer total sequences, and these dominant variant and high-frequency sequences were predicted to belong to individual PEs by Ecotype Simulation depending on the criteria used for demarcation. To simplify figures, all in-text figures report predominant PE populations representing ≥8% of the total diversity in an individual habitat sample; results for all PE populations are included in Appendix B.

Ecotype Simulation analysis of high-frequency sequences from the clade of A-like and A′-like lineages predicted 22 PEs (15 and 7 PEs, respectively) and analysis of the B′-like lineage predicted 24 PEs (Figure 3.1; also see Tables 3.1 and 3.2). PEs predicted from Ti454-barcode sequences (2008 collections) largely matched PEs previously demarcated in cloning and sequencing studies (2006 collections; Becraft et al, 2011). Newly discovered PEs in pyrosequencing analyses were labeled sequentially following demarcated in cloning and sequencing studies (2006 collections; Becraft et al, 2011). Newly discovered PEs in pyrosequencing analyses were labeled sequentially following the PEs previously identified and numbered in Becraft et al. (2011). Ecotype Simulation calculated periodic selection and ecotype formation rates of 2.71 (sigma) and 0.94 (omega) events/nucleotide substitution for A-like and A′-like sequences, and 1.86 and

0.94 events/nucleotide substitution for B′-like sequences, indicating that the rate of periodic selection was higher than the ecotype formation rate.

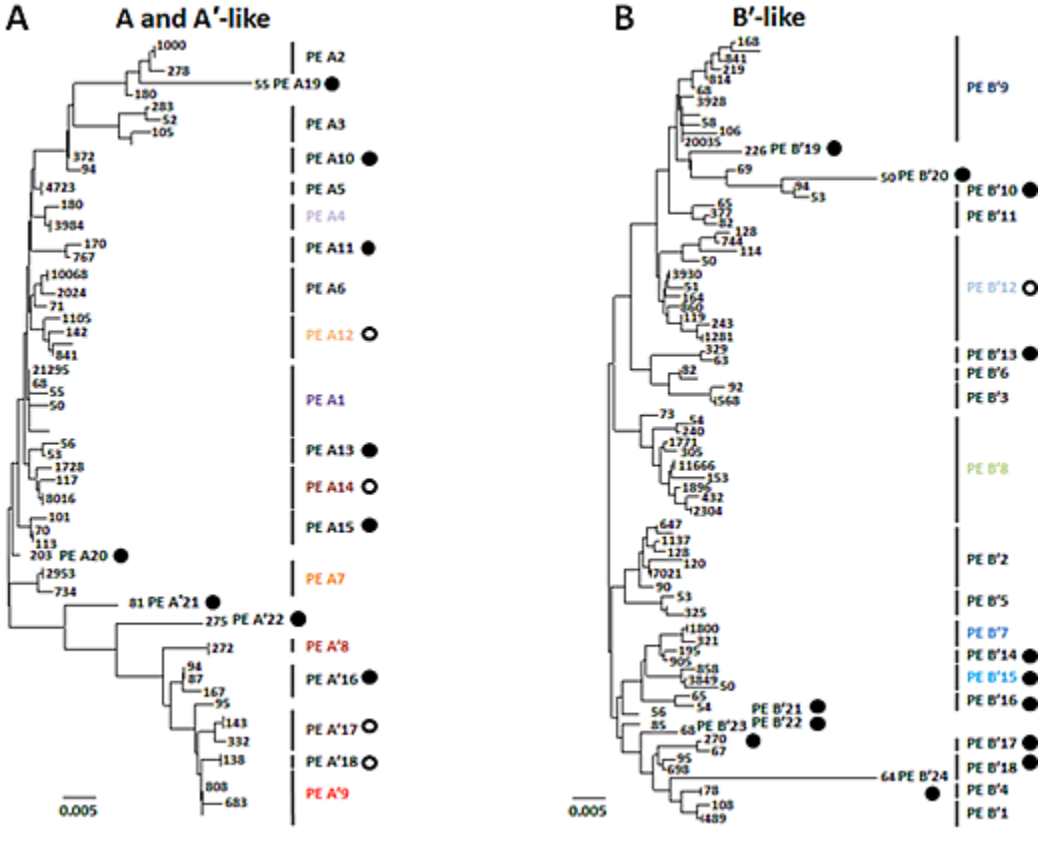


Figure 3.1. Neighbor-joining trees containing all high-frequency sequences (including dominant variants) identified by the Perl algorithm high-frequency sequence Finder. Black vertical bars represent putative ecotypes (PEs) predicted by Ecotype Simulation (fine-scale demarcation for A plus A′-like sequences, conservative demarcation for B′-like sequences). Numbers indicate the total number of sequences sampled with a particular variant across all 96 DNA samples. Filled circles represent PEs newly identified in Ti454-barcode analysis, and open circles indicate previously identified ecotype subclades in cloning and sequencing studies now demarcated as separate PEs. Colored PE designations represent predominant populations discussed in the main text, and colors correspond across all figures. Scale bars represent (A) 0.005 and (B) 0.002 substitutions per site.

Distribution Along the Effluent Flow Path

PEs were found in different relative abundances at different temperatures ($P<0.001$; see Appendix Table B1). While distributions overlapped across temperature sites, several abundant PEs declined in relative abundance abruptly and together at certain boundaries, thus demarcating three temperature-defined ecological zones (EZs), which we term EZ1 between ≤60 and 63°C, EZ2 between 64 and 65°C and EZ3 at ≥66°C, separated by two ecotones (between 63 and 64°C and between 65 and 66°C) (Figure 3.2).

Distribution Along Vertical Profiles

Vertical profiles in EZ1 showed clear evidence of differences in vertical distribution of PEs (Figure 3.3) ($P<0.001$; see Appendix Table B1). From the top to the bottom of the green cyanobacterial layer a progression was observed of PEs B′7 and B′9, to PE A1, then to a set of PEs including A4, A6 and A14. Additional profiles were assigned to EZs and are shown in Appendix Figure B3 and Appendix Tables B5 through C11, where it can be seen that despite sample to sample variation, trends in vertical distributions in EZ1 are evident. Generally, multiple ecotypes did not abruptly disappear together at certain boundaries, but coexisted along depth ranges. The only evidence of ecotones in the vertical distributions was the co-distribution of PEs A4, A6 and A14 deep in the EZ1 mat. These vertical distribution patterns alone do not distinguish among these three PEs, and could not be taken as evidence that these PEs are different ecological species. Since there are dramatic changes in the light and oxygen microenvironments in these mats (Figure 3.3B), as well as in pH (Jensen et al, 2011), PEs may be differently

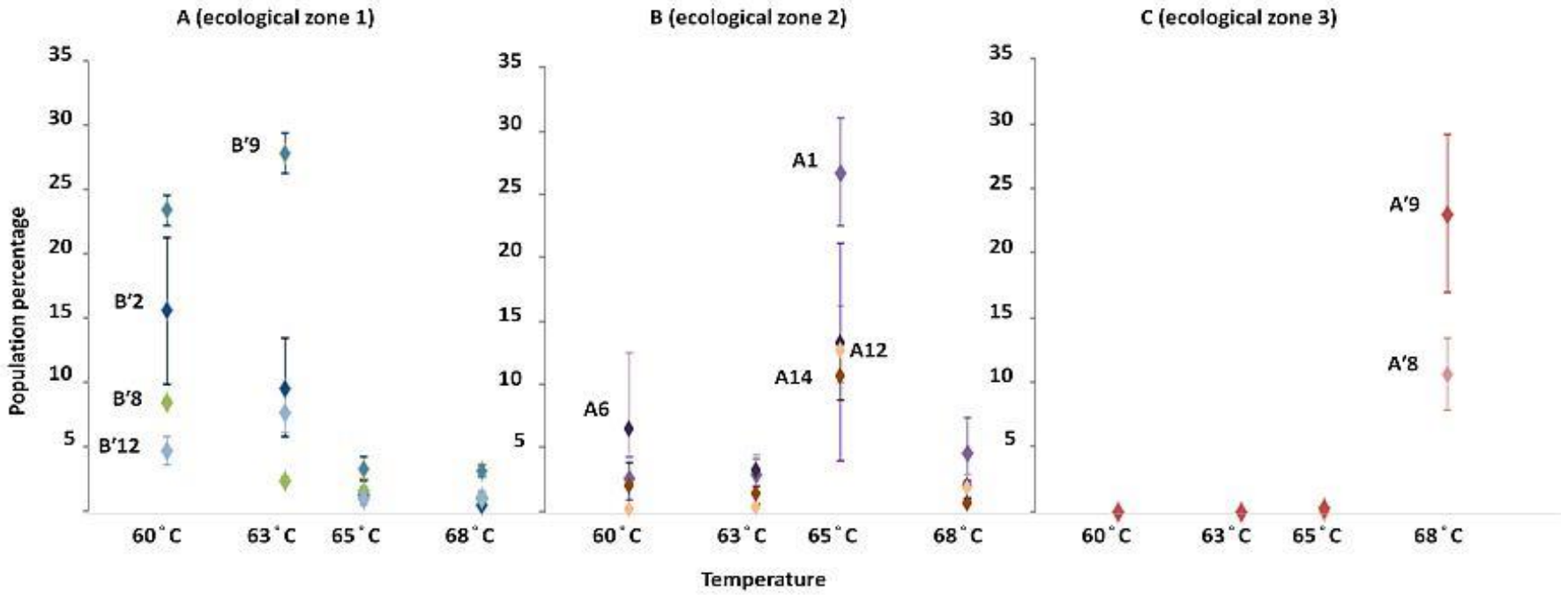


Figure 3.2. Relative abundance at four temperature sites along the effluent flow path of abundant putative ecotypes (PEs) associated with (A) ecological zone 1; (B) ecological zone 2; and (C) ecological zone 3. Samples correspond to average midday temperatures within ecological zones. Bars indicate the range between replicate samples; points indicate averages.

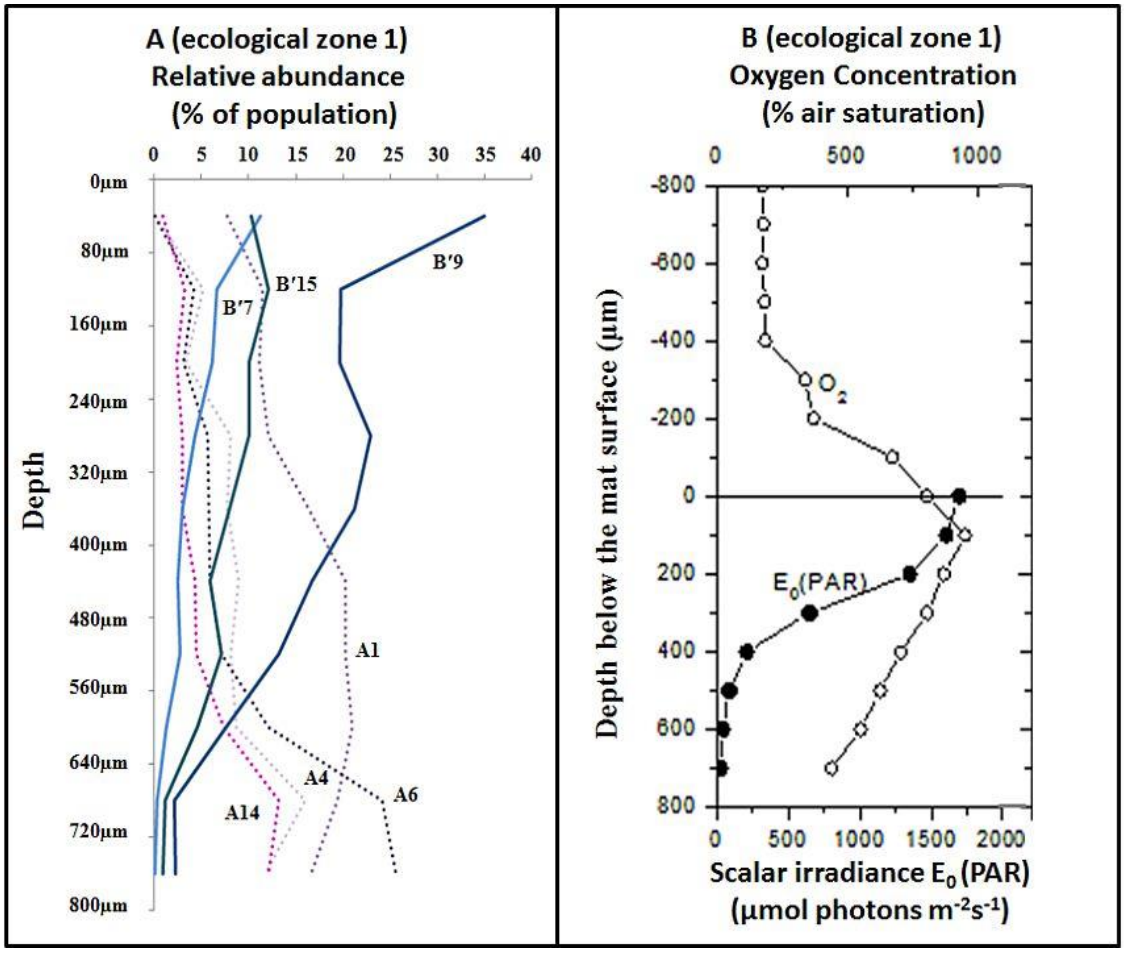


Figure 3.3**.** Relative abundance of predominant putative ecotypes in ~80 µm subsections along the vertical gradient (A) relative to light (PAR) and oxygen profiles (B) in ecological zone 1. Solid lines represent B′-like PEs and dashed lines represent A-like PEs. Additional profiles are shown in Supplementary Figure 3.

adapted to light intensity and/or quality or to other differences, such as availability of dissolved inorganic carbon.

psaA Transcription Patterns

In EZ1, *Synechococcus psaA* transcripts were expressed in a diurnal pattern (Figure 3.4A). High-throughput sequencing demonstrated differential expression of

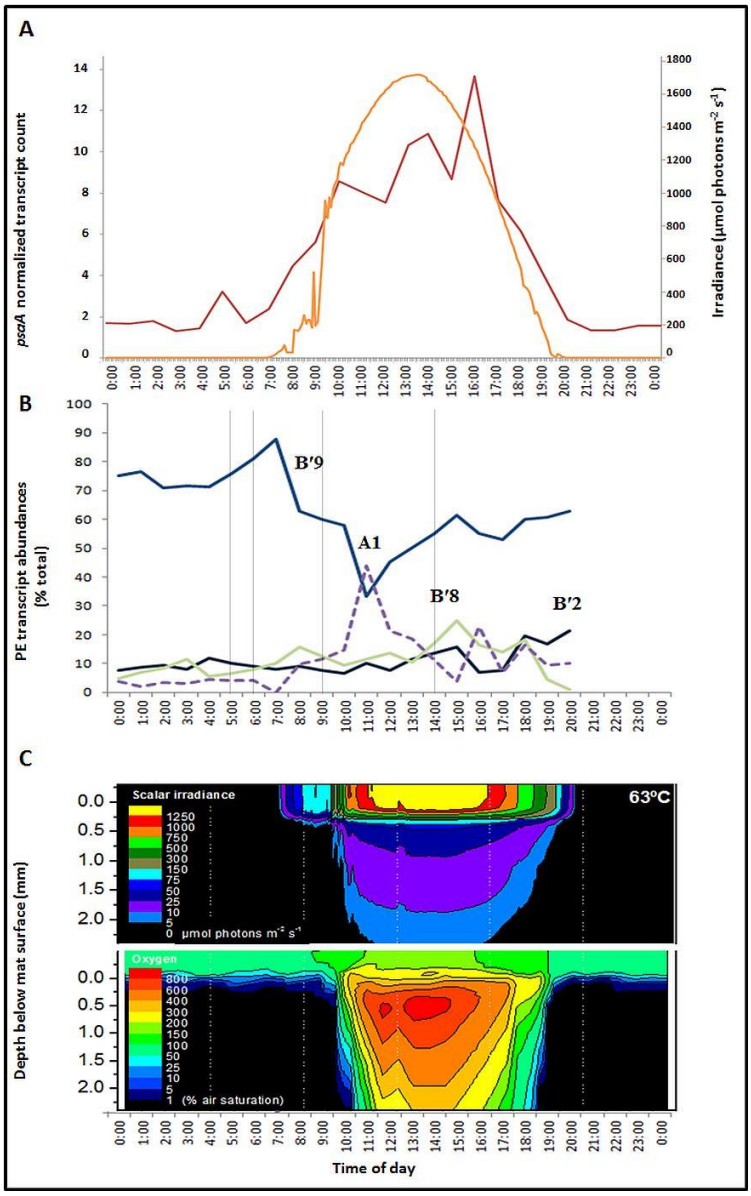


Figure 3.4. Diel expression of *psaA* transcripts. (A) Downwelling photon irradiance (400-700 nm) measured over a diel cycle on September 11 and 12 2009 (orange line), and normalized total *psaA* transcript count (red line). (B) Relative abundances of transcripts of abundant PEs in EZ1 measured from *psaA* cDNA over the diel cycle. Samples corresponding to time points at 0500, 0600, 0900, 1400 (vertical lines), 2100, 2200 and 2300 are missing due to failed sequencing reactions. Relative transcript abundances are high throughout the night most likely because of PCR amplification of low abundance sequences and expressing results as relative percent. (C) Isopleth diagrams showing depth distribution in EZ1 of scalar irradiance (µmol photons m^{-2} s^{-1}) and O_2 concentration (% air saturation) over the diel cycle. Samples at time points 1700, 1800, 1900 and 2000 h were from the previous evening, 11 September 2009.

transcripts contributed by the dominant *Synechococcus* PEs B′9 and A1, as well as the less abundant PEs B′2 and B′8 (Figure 4B) ($P<0.001$; Appendix Table B1). The later transcription of PE A1 transcripts may result from the limited penetration of light into the mat until between 1000 and 1100 h, when photon irradiance increased to 5-100 µmol photons m^{-2} s^{-1} between 320-720 µm in depth (Figure 3.4C), where this population is predominantly located (Figure 3.3A). Oxygen concentrations at these mat depths increased between 1000 h and 1100 h to 300-600 times air saturation, and further accumulated to >800 times air saturation by 1300 h, which reflects increased photosynthetic activity at depth (Figure 3.4C). In contrast, photon irradiance increased to between ~500-1250 µmol photons m^{-2} s^{-1} at the mat surface between 1000 h and 1100 h in parallel with a decline in PE B'9 transcript relative abundance. Lower relative abundance of *psaA* transcripts associated with PEs residing close to the mat surface under high irradiance might be the result of dilution by the increase in PE A1 transcripts, but results from a similar DGGE analyses (Appendix Figure B4) suggested that there was likely a mid-day decline in the abundance of PE B′9 transcripts.

PE-specific expression patterns over morning and evening light transition periods in EZ1, EZ2 and EZ3 are included in the Appendix (Appendix Figure B5 and Appendix Table B12). Those for EZ1 were similar to the results of the full diel pattern shown in Figure 3.4B, and a similar pattern was observed for A'-like PEs in EZ3. However, in EZ2, PEA1 transcripts were high at all times, though PE A1 was evenly distributed throughout the vertical aspect of the mat in EZ2 (Appendix Figure B3 E).

Responses to Environmental Perturbations

Environmental perturbation studies were conducted in order to test whether differences in PE distributions reflect adaptations to different temperature and irradiance levels. Results are summarized in Table 1. Data presented in Appendix Figure B6 demonstrate that during similar light alteration experiments in 1996 populations in control samples remained stable over the period of time studied ($p = 0.35$; see Appendix Table B2).

Temperature Increase. PE composition changed in the four days following shift of samples collected within EZ1 (60˚C) to a temperature characteristic of EZ2 (65˚C) (Figure 3.5A; Appendix Table B13), where the six most abundant ecotypes in the temperature manipulation experiment were significantly heterogeneous in their responses to increased temperature ($P=0.0155$; see Appendix Table B2). In particular, PEs A6, B′2, B′8 and B′9 declined, while PEs A1 and A14 increased in relative abundance. Effectively, the composition of abundant *Synechococcus* PEs within the sample shifted from that roughly characteristic of EZ1 to that roughly characteristic of EZ2, as expected from temperature distributions (Figure 3.2).

Light Alteration. Light alteration experiments were conducted at a site typical of EZ2 in order to focus on very closely related A-like PEs (also see Appendix Section B IV and Appendix Figure B7 A). The four most abundant putative ecotypes were significantly different in their responses to shading ($P=0.0262$; see Appendix Table B2).

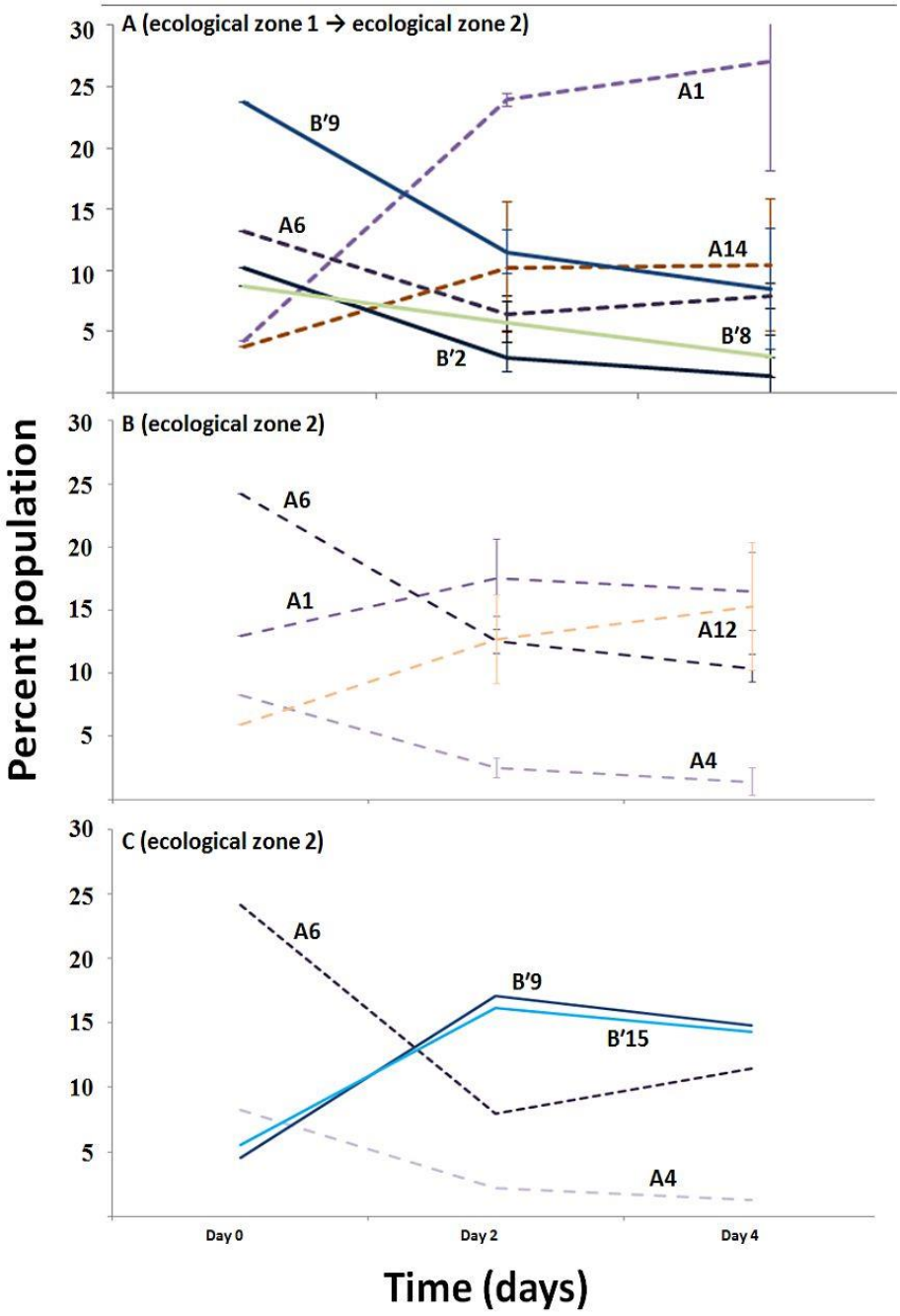


Figure 3.5. Change in relative abundance of abundant PEs over a four day period after environmental perturbations. (A) Temperature shift of mat samples from EZ1 to EZ2. (B) Reduction of downwelling photon irradiance by ~92% at EZ2. (C) UV light removal at EZ2. Solid lines represent B′-like PEs and dashed lines represent A-like PEs. Bars in A and B indicate the range between replicate samples.

Following removal of ~92% of the ambient light, PEs A4 and A6 declined, whereas PEs A1 and A12 increased in relative abundance (Figure 3.5B; Appendix Table B14). In response to nearly complete removal of UV light, PEs showed differential responses ($p < 0.001$; see Appendix Table B1). PEs B′15 and B′9, which are low-abundance PEs in EZ2, tripled in relative abundance (Figure 3.5C). PEs A4 and A6 both substantially decreased in relative abundance, responding in a manner similar to that observed in the light reduction experiment. Unfortunately, replicate samples from the UV light exclusion study were lost during high-throughput analysis, however, the differential response of surface-oriented PE B′9 in a lower temperature mat sample (EZ1) was observed in another UV blocking experiment (see Appendix Figure B7 B).

Genetics of Population Response. Here we tested whether the variants contained within a given PE were really ecologically homogeneous as predicted. We attempted to reveal ecological variation within a PE by determining whether high-frequency sequences within a PE changed differently, and if the ratio of all low-frequency sequences within a PE to the total count of sequences within the PE ((low-frequency sequence/ (dominant variant + high-frequency sequence+ low-frequency sequence)) x100) remained the same as PE relative abundances shifted after an environmental perturbation (Figure 3.6). Throughout the light reduction experiment high-frequency sequences changed in concert with the dominant variant sequence, and the ratio of low-frequency sequence variants to the total sequences of abundant PEs never fluctuated more than 6% between samples ($P>0.25$; see Appendix Table B3), despite increases or decreases in PE relative

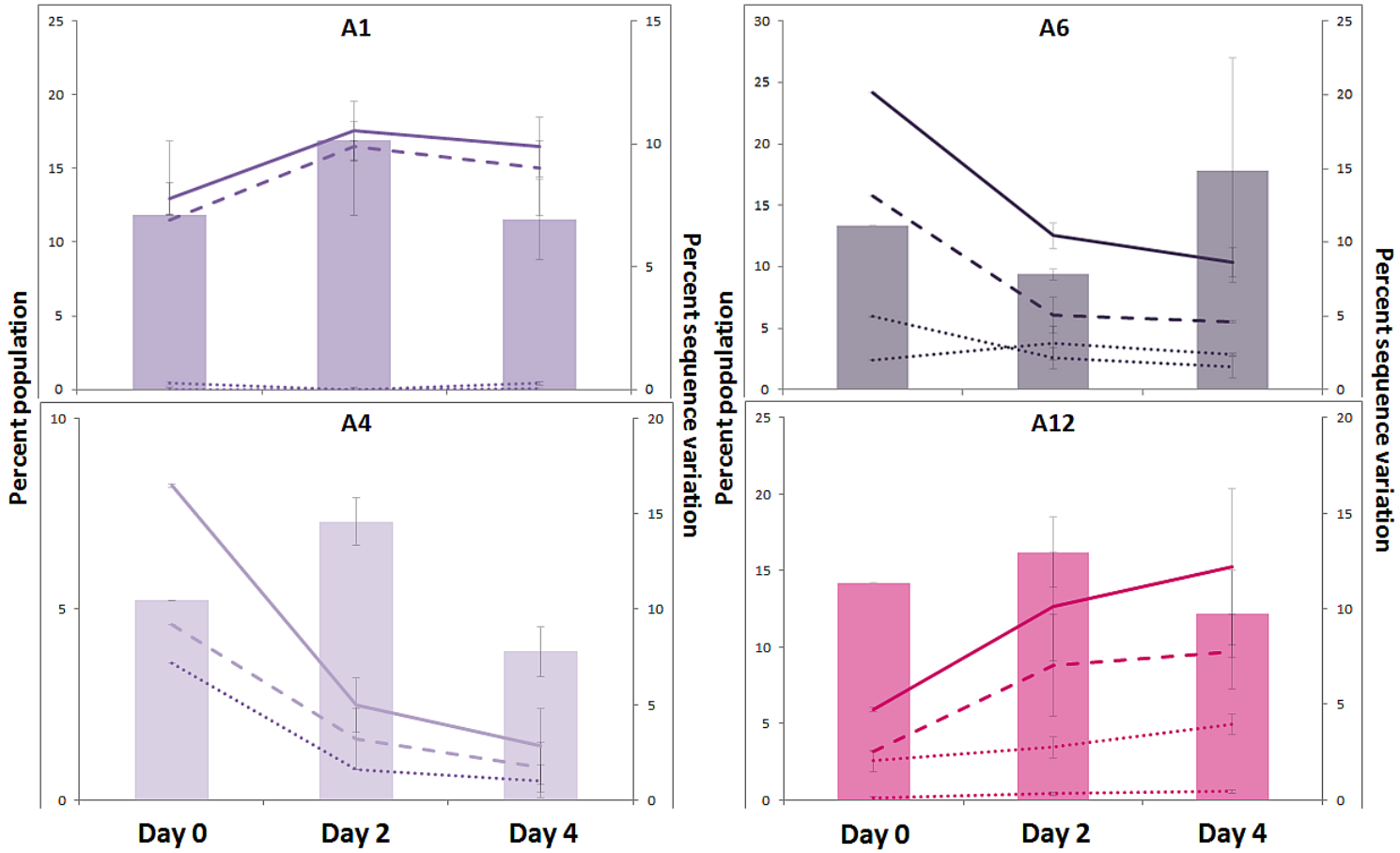


Figure 3.6. Percent in population of dominant variant sequences (solid lines; left axis), high-frequency sequences (dashed lines; left axis) and total low-frequency sequence variants (bars; right axis) in putative ecotypes (PEs) undergoing change in relative abundance over a four day period in ecological zone 2 after reducing downwelling photon irradiance by ~92%. Total variation within PEs is represented by solid lines. Bars indicate the range between replicate samples.

abundances in response to environmental perturbation. Similar results were observed in temperature-shift experiment (Appendix Figure B8). That the membership of a given PE (dominant variant, high-frequency sequences and low-frequency sequences) rose or fell together in response to environmental change indicates that the individuals contained within a PE were ecologically interchangeable.

Discussion

The deeper coverage of genetic variants and habitats provided by pyrosequence analyses resulted in the demarcation by Ecotype Simulation of a greater number of PEs (15 A-like, 7 A'-like and 24 B'-like PEs) than had been observed in earlier cloning and sequencing studies (10 A-like, 2 A′-like and 20 B′-like PEs) (Becraft et al, 2011). Although more samples were included in the barcode study, all high-frequency sequences were detected in habitat samples that corresponded to the cloning and sequencing work, suggesting that the greater number of PEs demarcated resulted from the deeper coverage provided by high-throughput sequencing. PEs that were newly predicted in barcode analyses corresponded to subclades previously embedded within PEs that had not been identified from more limited clone sequence data (e.g., PE A12; as shown in Tables 1 and 2), or to low-abundance populations not detected in clone libraries (e.g., PE A10). Cloning and sequencing studies analyzed 0.1% of the amount of sequence data obtained in the 2008 Ti454-barcode study, yet detected 70% of the currently demarcated PEs. In addition, newly demarcated PEs were in low relative abundance, indicating that the predominant *Synechococcus* populations in these habitats have been identified.

Differences in the relative abundances of PEs suggested minor changes, perhaps due to environmental change or random drift in species abundance. For instance, PEs B′10, B′11 and B′12, which were detected in low abundance in 2006, were not detected in 2008. Also, PEs B′3 and B′5 dominant variants and high-frequency sequences were less frequently detected in 2008 than in 2006. The deep sequencing provided by barcode analyses also allowed us to detect subtle genetic differences within PEs. The genetic structure of PE populations remained relatively constant, although minor differences were observed, indicating that the Mushroom Spring *Synechococcus* populations were relatively stable throughout the study period.

As summarized in Tables 3.1 and 3.2, Ti454-barcode technology allowed for the confirmation of ecological distinction for most of the predominant PEs predicted by Ecotype Simulation. PEs were shown to have different distributions along flow and vertical gradients at given temperatures, different ecotype-specific expression patterns that corresponded to their vertical distributions in the mat, and different responses to environmental perturbations. The distribution of PEs within samples collected along a temperature gradient in 2008 and analyzed by Ti454-barcoding corresponded to EZs suggested by analysis of samples collected in 2006 and analyzed by cloning and sequencing (Tables 1 and 2; Appendix Table C4). Diversity observed at the *psaA* locus between ecotones correlated to 16S rRNA sequence diversity along the flow path observed in Ferris and Ward (1997), where B′-like 16S rRNA diversity was detected between ~57 to 63°C, A-like diversity was detected between ~60 and 65°C, and A′-like

Table 3.1. Summary of abundant A and A′-like PE population percentages for individual habitat samples corresponding to distributions, expressing timing and responses to environmental perturbations. PEs without identified vertical distributions or perturbation responses (arrows) were too low in relative abundance to track with confidence.

		Percent population and vertical position											Expression timing			Perturbation response		
		2006 cloning & sequencing					2008/09 Ti454 bar code											
PE Ti454	Becraft et al. 2011 designation [1]	60°C (EZ1)	2 V	63°C (EZ1)	V	65°C (EZ2)	60°C (EZ1)	V	63°C (EZ1)	V	65°C (EZ2)	V	68°C (EZ3)	63°C (EZ1)	65°C (EZ2)	68°C (EZ3)	60→65°C (EZ1→2)	light reduction (EZ2)	UV reduction (EZ2)
A1	A1 (A1-3, A1-4)	9.1	↓	43.8	↕	6.2	2.6	↓	2.9	↓	26.8	↕	4.7	0930-1300	0700-1900		↑	↑	-
A2	A2	0.4		0.0		3.3	0.2		0.3		0.4		2.0				-	↓	↓
A3	A3	1.1		0.0		0.0	0.1		0.1		0.5		0.0				-	-	-
A4	A4	4.2		0.0		0.0	2.7	↓	0.5		3.6	↓	0.6				↑	↓	↓
A5	A5	1.5		0.0		0.0	1.5		1.0		8.1	↓	0.6				↑	↑	↓
A6	A6	3.8		5.6		1.6	6.4	↓	3.2	↓	13.2	↔	1.9	0930-1100	0800 and 1900		↓	↓	↓
A7	A7	1.1		50.0	↑	1.6	0.0		0.2		4.5	↕	3.2		0700-1100		-	↑	↓
A′8	A′8	0.0		0.0		18.0	0.0		0.0		0.0		10.7			1900	-	-	-
A′9	A′9 (A′9-2)	0.0		0.0		52.5	0.0		0.0		0.3		23.1			1100	-	-	-
A10							0.3		0.2		0.8		0.5				-	↓	-
A11							0.1		0.1		1.4		0.1				-	↓	↓
A12	A1-2	0.0		16.7		0.0	0.1		0.3		12.6	↕	1.8				-	↑	↑
A13							0.0		0.0		0.1		0.2				-	-	-
A14	A1-1	3.8		0.0		0.0	2.0	↓	1.3	↓	10.7	↔	0.6				↑	↑	↓
A15							0.1		0.0		0.7		0.1		0700-1100		↑	-	-
A′16							0.0		0.1		0.2		8.3			1300	-	-	-
A′17	A′9-1	0.0		0.0		11.5	0.0		0.0		0.0		14.6			1900	-	-	-
A′18							0.0		0.0		1.2		0.4				-	-	-

1 PE subclades that were demarcated as a single ecotype in Becraft et al. (2011) that were separately demarcated in barcode analyses are indicated with hyphenated numbers next to PE designations. PE subclades that were not detected in barcode analysis due to decreased sequence length are indicated in parentheses next to the PE they were demarcated with in Becraft et al. (2011).

2 ‘V’ indicates vertical positioning at a specified temperature. Arrows under ‘V’ indicate highest relative abundance in upper mat layers (↑), highest relative abundance in lower mat layers (↓), highest relative abundance in center layers (↔), and relative abundance through all layers (↕).Perturbation responses are indicated by ↑ for increase in relative abundance, ↓ for decrease, and - for no change.

Table 3.2. Summary of abundant B′-like PE population percentages for individual habitat samples corresponding to distributions, expression timing and responses to environmental perturbations. PEs without identified vertical distributions or perturbation responses (arrows) were too low in relative abundance to track with confidence.

		Percent population and vertical position											Expression timing			Perturbation response		
		2006 cloning & sequencing					2008/09 Ti454 bar code											
PE Ti454	Becraft et al. 2011 designation 1	60°C (EZ1)	2 V	63°C (EZ1)	V	65°C (EZ2)	60°C (EZ1)	V	63°C (EZ1)	V	65°C (EZ2)	V	68°C (EZ3)	60°C (EZ1)	65°C (EZ1)	68°C (EZ1)	60→65°C (EZ1→2)	light reduction (EZ2)	UV reduction (EZ2)
B′1	B′1	6.9		0.0		0.0	0.4		0.3		0.0		0.4				-	-	↑
B′2	B′2	3.1		0.0		0.0	15.7	↓	9.7		1.4		0.5	0800 and 1300			↓	-	↑
B′3	B′3	2.3		0.0		0.0	0.1		0.3		0.0		0.9				-	-	-
B′4	B′4	4.2		0.0		0.0	0.0		0.2		0.0		0.0				-	-	-
B′5	B′5	1.1		0.0		0.0	0.0		0.3		0.1		0.0				-	-	-
B′6	B′6	1.1		0.0		0.0	0.1		0.0		0.0		0.8				-	-	-
B′7	B′7	8.0	↑	0.0		0.0	0.8		5.2	↑	0.1		0.2				-	-	↑
B′8	B′8	8.4		0.0		0.0	8.6	↑	2.5		1.6		1.0	1100-1300			↓	-	↑
B′9	B′9	29.9	↑	0.0	↑	0.0	23.5	↑	27.9	↑	3.4		3.2	1100-1300			↓	↓	↑
none	B′10	1.1		0.0		0.0													
none	B′11	1.1		0.0		0.0													
none	B′12	0.8		0.0		0.0													
B′10							0.2		0.0		0.2		0.0				-	-	-
B′11							0.6		1.0		0.1		0.1				-	-	-
B′12	B′9-1 and B′9-2	11.1		0.0		0.0	4.8	↑	7.8		1.0		1.1				↓	-	↑
B′13							0.1		0.2		0.0		0.1				-	-	-
B′14							0.0		1.0		0.0		0.1				-	↑	↑
B′15							0.2		0.3	↑	5.1		1.0				↑	-	-
B′16							0.0		0.4		0.0		0.0				-	-	-
B′17							0.9		0.8		0.0		0.0				-	-	-
B′18							0.8		0.7		0.1		0.0				-	-	-

1 PE subclades that were demarcated as a single ecotype in Becraft et al. (2011) that were separately demarcated in barcode analyses are indicated with hyphenated numbers next to PE designations.

2 ‘V’ indicates vertical positioning at a specified temperature. Arrows under ‘V’ indicate highest relative abundance in upper mat layers (↑), highest relative abundance in lower mat layers (↓), highest relative abundance in center layers (↔), and relative abundance through all layers (↕). Perturbation responses are indicated by ↑ for increase in relative abundance, ↓ for decrease, and - for no change.

diversity was detected at >65°C. Temperature shift experiments provided further evidence that different PEs exhibit different temperature preferences, consistent with results of previous culture studies (Miller and Castenholz, 2000; Allewalt et al, 2006).

Vertical distribution results also matched patterns presented in Becraft et al. (2011), where bands that corresponded to dominant variants observed in DGGE gels often showed significant overlap with other dominant variant bands, even though band intensity varied, and was most intense within a particular depth range. Differential PE responses to reduction in light intensity suggest that adaptation to light may, at least in part, underlie these distribution patterns. Indeed, in other ongoing studies we have observed that strains representative of populations residing at different depths have adaptations to light intensity that are consistent with the amount of light present at these depths (Nowack, S., M. Olsen, E.D. Becraft and D.M. Ward, unpublished). However, without quantification the relative responses to light alteration are difficult to interpret. While populations can be seen to respond differently, it is not known whether the decreasing relative abundance of one PE is due to the increasing relative abundance of another PE, or vice versa.

Differential responses of PEs to UV light removal expanded upon the results of Miller et al. (1998), who demonstrated that photosynthesis in suspensions of mat samples was different with full-light exposure than when UV light was removed. In Norris et al. (2002), *Synechococcus* 16S rRNA genotypes showed little change after a two month period of exclusion of UV light. PEs predicted from the more highly resolving *psaA* locus

showed population changes after only two days of UV exclusion, suggesting that the 16S rRNA locus is too conserved to accurately demarcate ecologically distinct species.

As has been previously observed for other photosystem reaction center genes (Steunou et al, 2008; Jensen et al, 2011; Liu et al, 2012), *psaA* transcripts showed a diurnal expression pattern. The earlier expression of PE B'9 transcripts compared to PE A1 transcripts matches the relatively longer and shorter photoperiods experienced by these surface- and subsurface-associated populations, respectively. Interestingly, Ramsing et al., (2000) conducted microsensor studies that demonstrated that peak oxygenic photosynthesis moved downward to deeper mat layers as light intensity rose during the morning (Ramsing et al., 2000). The transcription results are also consistent with the downregulation of photosystem I in cyanobacteria under high-light conditions (Muramatsu et al, 2009). Interestingly, the surface-associated PE B'9 population consistently increased in relative abundance following removal of UV light, possibly suggesting its sensitivity to high UV irradiance. UV sensitivity of a surface *Synechococcus* species population might explain the decline in photosynthesis in the afternoon observed by Miller et al. (1998).

The high-throughput analysis of sequence variation of an essential divergent protein-encoding locus showed that most relatively abundant PEs predicted through evolutionary simulation analysis meet the expectations of ecological species. They are ecologically distinct populations and, importantly, the genetic diversity within PEs is maintained as they increase or decrease in relative abundance in response to environmental perturbations. Diversity within populations changing in unison in response

to environmental perturbation indicates the ecological interchangeability of the individuals within those populations, and is an expected property of ecological species. These and other results (Melendrez et al., 2011) indicate that molecular divergence among ecological species is much less than among the species recognized by systematists, including definitions of 1-3% 16S rRNA divergence (Wayne et al, 1987; Stackenbrandt and Ebers, 2006) or >95% average nucleotide identity (ANI) (Konstantinidis and Tiedje, 2005; Konstantinidis, 2011). In the case of hot spring *Synechococcus* spp. we have observed evidence that temperature adaptation, which can be observed in 16S rRNA distribution studies (Ferris and Ward, 1997), most likely preceded light adaptation, which must be observed using more rapidly evolving genetic markers. Interestingly, while the problem of molecular resolution is similar, in the case of marine *Prochlorococcus* light adaptation appears to be more ancient than temperature adaptation (Martiny et al. 2009). In these and other habitats, where microbial diversity is organized by environmental parameters, ecological species populations are likely to be the fundamental units of community composition and structure.

Acknowledgements

This research was supported by the National Science Foundation Frontiers in Integrative Biology Research Program (EF-0328698), the National Aeronautics and Space Administration Exobiology Program (NNX09AM87G), the Danish Council for Independent Research | Natural Sciences (to M.K. and S.I.J.), and the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of

BER′s Genomic Science Program 395 (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL) under contract 112443. This study was conducted under Yellowstone National Park research permits YELL-0129 and 5494 (D.M.W.) and YELL-02058 (M.K.), and we appreciate the assistance from National Park Service personnel. We thank Jennifer Weeding from Montana State University for her help with the statistical analysis, and we thank Manolis Kaparakis of Wesleyan University for help with Stata. We also thank Jay Liu and Don Bryant for providing *psaA* transcript data for the diel study.

CHAPTER 4

DISTRIBUTION OF A/B-LINEAGE *SYNECHOCOCCUS* GENETIC AND ECOLOGICAL DIVERSITY IN HOT SPRINGS OF THE AMERICAN NORTHWEST

Introduction

Evolutionary processes, historical events and physiological constraints all dictate the content of the regional species pool, which can be thought of as the upper limit of the species composition for a given area (Figure 4.1) (Morin, 2011). Species usually cannot thrive outside their physiological adaptive range, limiting the local distribution to habitats with physiochemical parameters that can be tolerated. Dispersal capabilities and habitat selection (e.g. presence of needed resources) impact local population diversity by determining which species can arrive at and colonize a given location, essentially filtering species from the region. Finally, interspecies interactions affect the success or failure of any species within or entering the local community. While the local microbial communities inhabiting the mats of Octopus Spring and Mushroom Spring are becoming increasingly well understood, the relations of the species in these local communities to the regional species pool have not been as thoroughly studied.

As predicted by the Stable Ecotype Model (Chapter 1, Figure 1.9, p.18), populations can diverge though local niche adaptation as a consequence of ecological specialization to varying environmental parameters (sympatric speciation). Alternatively, the Geotype Plus Boeing Model (Figure 4.2) predicts that populations can also diverge

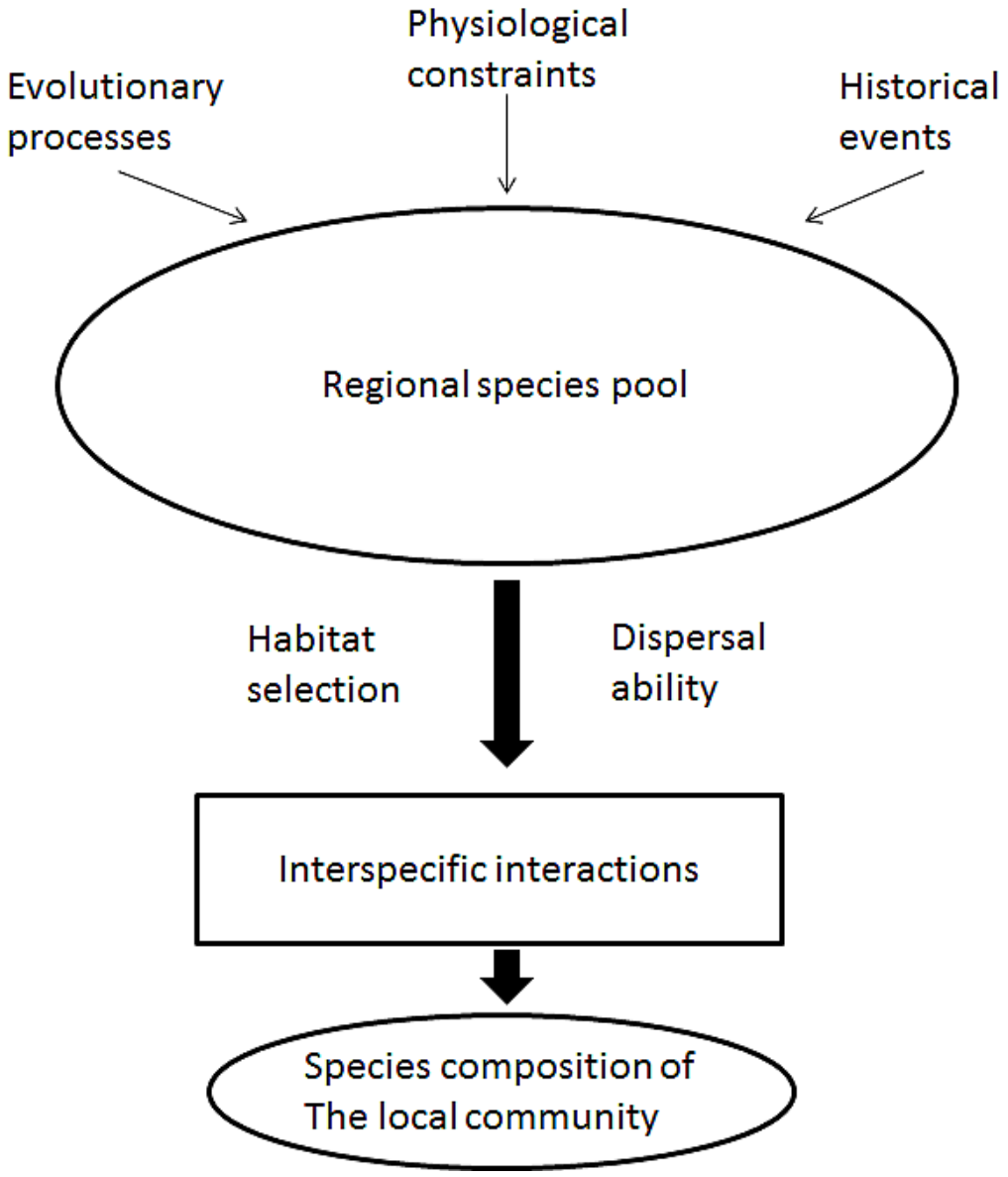


Figure 4.1. Relationship of species in a local community to the regional species pool. The composition of the species pool of potential community members depends on past evolutionary and historical events, as well as physiological constraints. Dispersal ability and habitat selection influence which members of which species can manage to arrive in a particular location. Interspecific interactions among those species that manage to arrive in a particular place further inhibit or facilitate the inclusion of species in the community (from Morin, 2011).

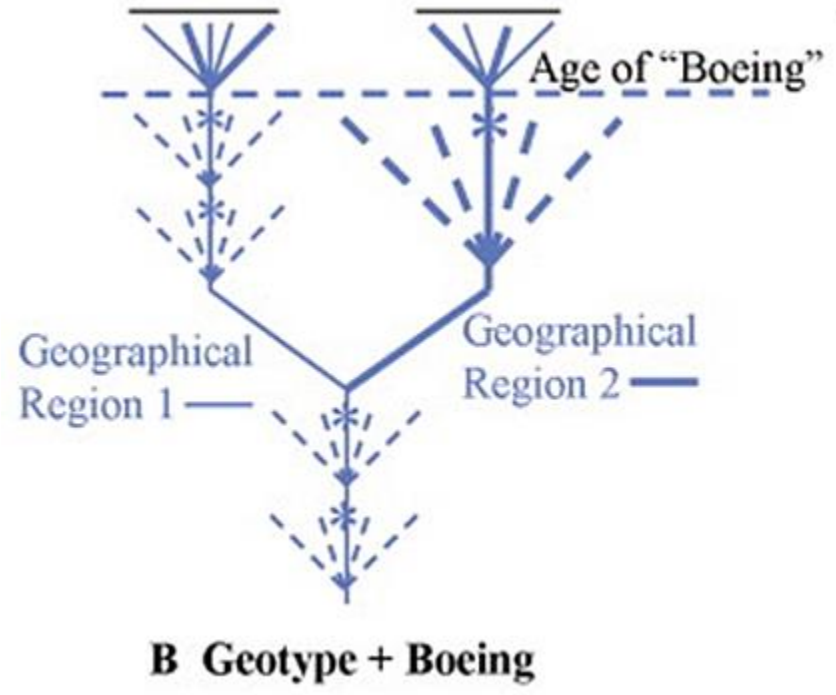


Figure 4.2. The Geotype + Boeing model of species and speciation. A history of geographic isolation, followed by recent human-aided transport (during the 'age of Boeing'), can lead to allopatric species with similar adaptations being separately demarcated (e.g. > 1 clade per ecotype). Lineages from different regions are represented by different line thicknesses (Cohan and Perry, 2007). Details are as described in Chapter 1, Figure 1.9, p.18, except that the single color indicates that the two clades are formed within a single ecotype.

through mutation and recombination if physical barriers, or separation by large geographical distances, reduce or prevent dispersal (allopatric speciation). In this case, two phylogenetic clusters (distinct geotypes) are formed within a single ecotype and, in contrast to the 1:1 ratio of clusters and ecological species in the Stable Ecotype Model, the ratio is 2:1 (or many to one). As the name of the model suggests, in the age of "Boeing", when long-distance travel is more frequent, the redistribution of geotypes could complicate the sequence-based demarcation of ecological species.

For many years it was widely believed that barriers to the distribution of microbial species did not exist, and that "everything is everywhere, and the environment selects" (Bass-Becking, 1934). In essence, microorganisms were thought to primarily evolve through sympatric means, and to be distributed globally without geographical barriers limiting their dispersal. Observations leading to such inferences were often based on organism morphology (Fenchel and Finlay, 2004). However, morphology is often a poor indicator of species richness and can mask the genetic diversity that exists in nature. Hot spring *Synechococcus* spp. provide a good example of this problem (see Chapters 1-3). It can be challenging to study the spatial dynamics of microbial populations at large-scale because (i) evolutionarily distinct organisms can share similar morphologies, (ii) cultured organisms are not always representative of predominant natural populations (Ward et al., 1990), (iii) different taxa can have a variety of evolved dispersal mechanisms (McDougald et al., 2012) and (iv) the immense bacterial diversity that exists in nature can make identifying ecological species in their natural habitats difficult (Dykhuizen, 1998; Ward et al., 2006).

Despite these complications, modern molecular technologies have allowed many researchers to show biogeographic patterns of microbial distributions (Hughes-Martiny et al., 2006). Pathogenic and symbiotic microorganisms that are associated with specific eukaryotic hosts have been shown to be restricted to the range of their hosts (Falush et al., 2003; Taylor et al., 2005), microbial diversity has been shown to increase with increased geographic area (Noguez et al., 2005; Reche et al., 2005; Fierer, 2008), geographic barriers were shown to isolate populations of hyperthermophilic Archaea (Whitaker et al.,

2003), and thermophilic cyanobacterial populations were shown to be differentiated by geography (Papke et al., 2003; Miller et al., 2006; Ward and Castenholz, 2000; Ward et al., 2012). These studies provided evidence that geographic isolation and allopatric speciation have been important factors in the evolution of microbial populations.

In Papke et al. (2003), the 16S rRNA locus and 16S-23S internal transcribed spacer (ITS) region were sequenced to determine whether hot spring cyanobacterial populations showed unique patterns of distribution globally and regionally. If hot springs act as isolated island-like environments, restricting gene flow between populations, and if this is not countered by frequent dispersal, such systems should facilitate evolutionary divergence between geographically separated microbial populations. Papke et al. (2003) collected hot spring mat samples from multiple springs within multiple geographic regions in the Northwest United States, Japan and New Zealand, so that they could study geographical variation on spatial scales ranging from springs within individual basins to springs on different continents. They made additional collections at some springs and basins to enable studies of local distributions along ecological gradients (e.g., temperature and pH). Springs in different countries are color-highlighted in Figures 4.3 and 4.4 and abbreviated names of all springs are provided in the Figure 4.4 legend.

Papke et al. (2003) showed that cyanobacterial 16S rRNA genotypes had different global distributions (Figure 4.3), even when inhabiting chemically similar hot springs (Figure 4.4 and Table 4.1A and B). The *Synechococcus* A/B lineage was shown to be endemic to North America, and the *Oscillatoria amphigranulata*-like lineage was

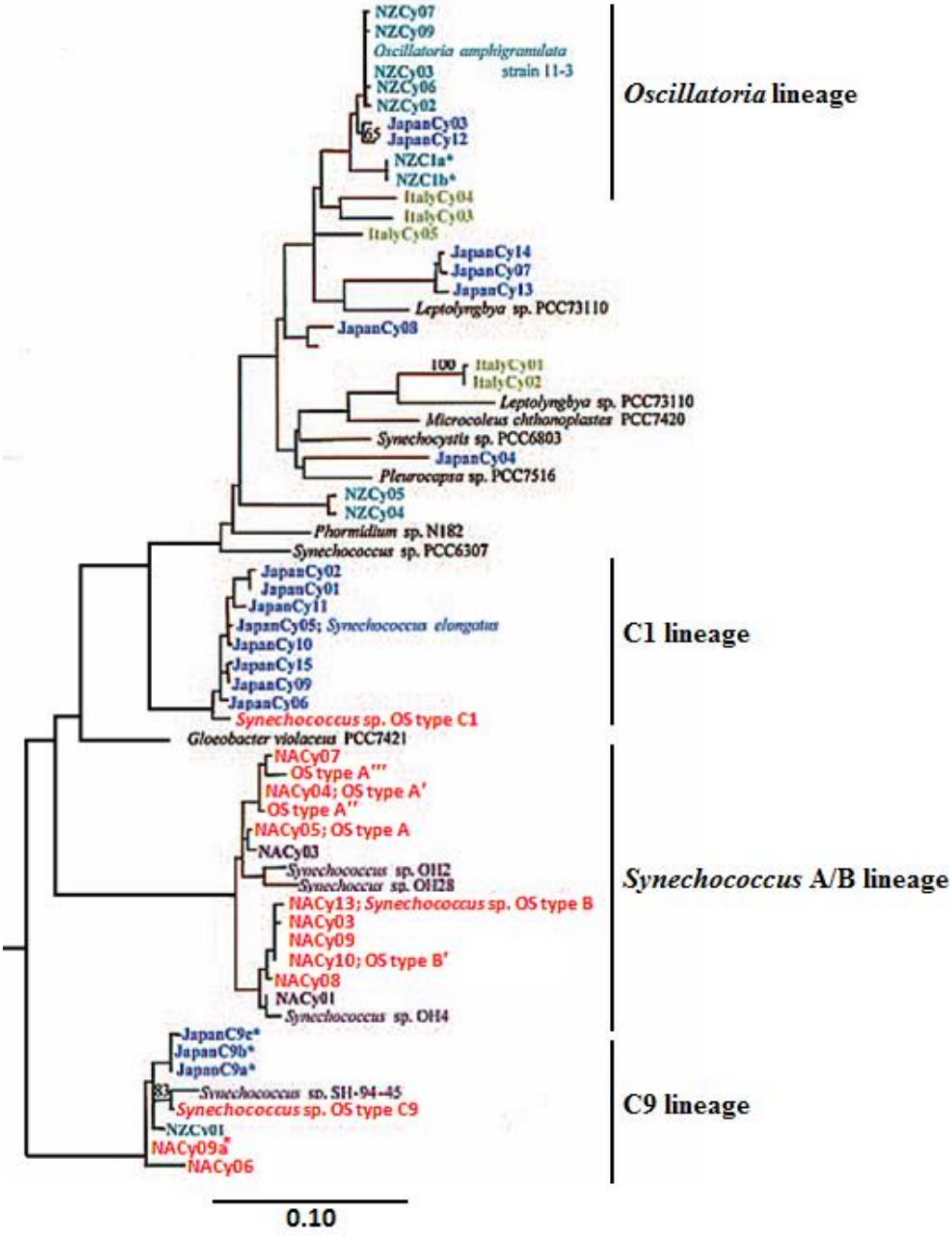


Figure 4.3. 16S rRNA phylogenetic tree demonstrating the relationships of cyanobacterial clones retrieved from hot spring cyanobacterial mats in different countries to other cyanobacterial 16S rRNA sequences, including hot spring *Synechococcus* spp. and *Oscillatoria amphigranulata* isolates. The tree was rooted with *Escherichia coli* and *Bacillus subtilis* 16S rRNA sequences. Color highlighting: green, New Zealand; blue, Japan; yellow, Italy; red, Greater Yellowstone Ecosystem; purple, Oregon. Scale bar represents 0.10 substitutions per site (modified from Papke et al., 2003).

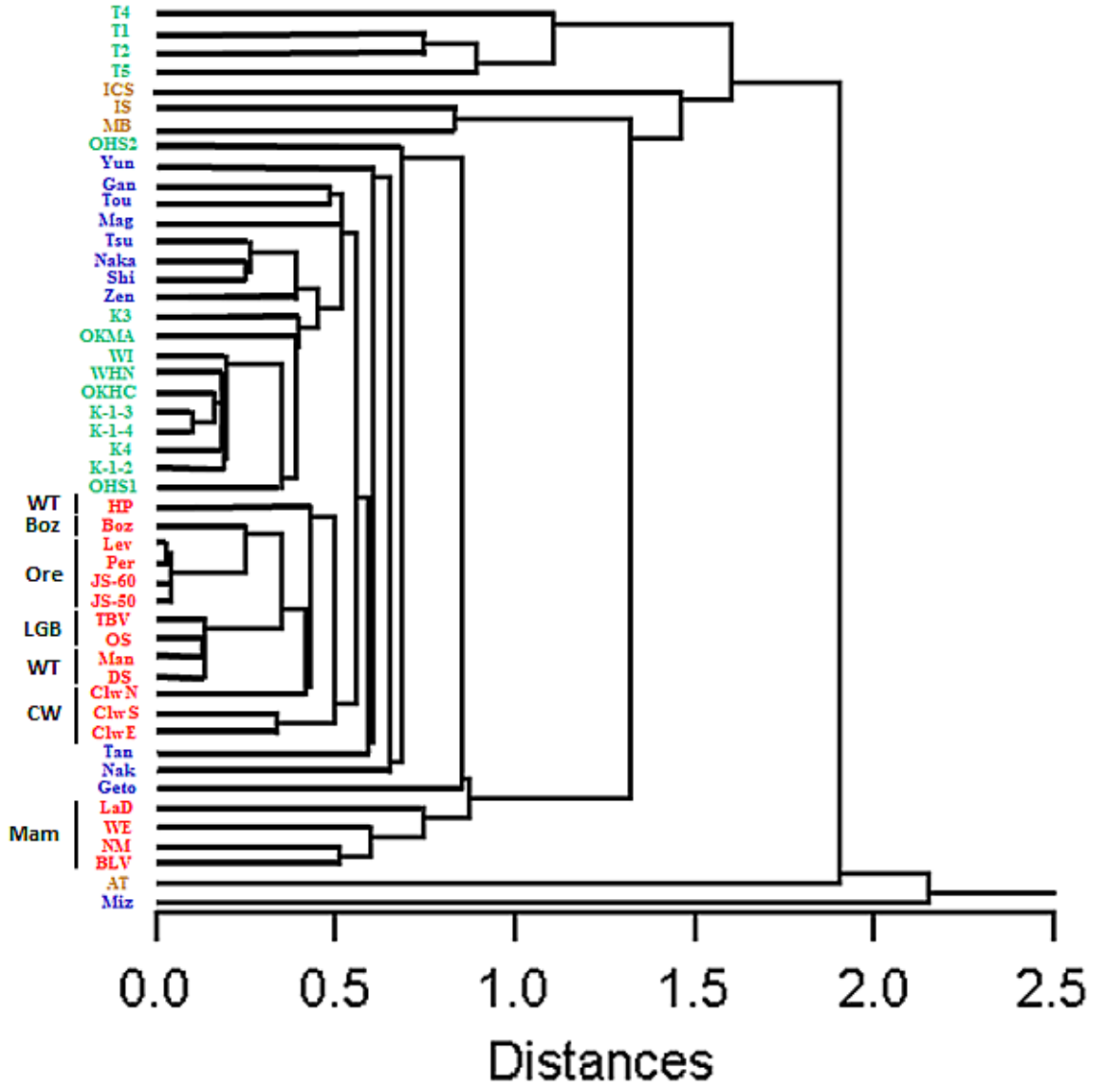


Figure 4.4. Hierarchical cluster analysis of chemical parameters in hot springs included in biogeographical analyses of Papke et al. (2003). Hot springs are color-coded by country: green, New Zealand; blue, Japan; yellow, Italy; and red, North America. JS, Jack Stream; Per, Perpetual; Lev, Levee; Boz, Bozeman; LaD, La Duke; NM, New Mound; BLV, Bath Lake Vista; WE, White Elephant Back; CWE, Clearwater East; CWS, Clearwater South; CWN, Clearwater North; OS, Octopus; TBV, Twin Butte Vista; MS, Mushroom; Man, Mantrap; HP, Heart Pool (modified from Papke et al., 2003).

Table 4.1A. Physical and chemical parameters for hot springs sampled across the North Western United States (modified from Papke et al., 2003).

Spring	°C	pH	PO_4[a]	NH_3[a]	NO_2[a]	NO_3[a]	$CaCO_3$[b]	Ca[b]	Cl[b]	Mg[b]
JS-60	59.8-61.2	8.1	0.37	2.5	0.57	0.54	64	9	130	0.5
JS-50	49.7-50.2	8.8	0.04	2.5	0.57	0.54	65	9	130	0.5
Per	60.1-60.9	8.2	0.37	4.1	0.57	0.54	65	10	120	0.5
Lev	59.5-61.8	8.1	0.37	10	0.57	0.54	72	10	120	0.5
Boz	57.2-57.5	8.8	0.45	2.5	0.57	0.54	87	1	53	0.5
LaD	59.8-60.2	6.7	0.37	2.5	1.09	0.54	230	340	49	59
NM	56.4	6.6	0.37	24	0.57	0.54	720	350	160	65
BLV	57.0-57.3	6.6	0.37	27	0.57	0.54	720	330	170	63
WE	59.7-60.0	6.9	0.37	19	0.57	0.54	610	330	170	67
CW E	57.3-61.9	5.2	0.37	14	0.57	0.54	8	6	130	0.5
CW S	56.3-56.5	6.1	0.37	6.5	0.57	0.54	31	7	140	0.5
OS	54.0-64.0	8.2	0.37	35	1.09	6.13	290	0.5	290	0.5
TBV	59.3-60.7	8.4	0.37	2.5	0.57	0.54	270	0.5	320	0.5
Man	58.2-61.8	8.3	0.37	2.5	0.57	1.12	350	0.5	180	0.5
HP	56.6-58.3	9.2	0.37	2.5	0.57	2.58	480	0.5	350	0.5

[a] µM.
[b] mg

Table 4.1B. Physical and chemical parameters for hot springs sampled across the North Western United States (modified from Papke et al., 2003).

Spring	°C	pH	SiO_2[b]	Na[b]	SO_4[b]	As[b]	B[b]	Cr[b]	Cu[b]	Fe[b]	Mn[b]	Zn[b]
JS-60	59.8-61.2	8.1	73	190	270	0.2	7.4	0.01	0.01	0	0	0.03
JS-50	49.7-50.2	8.8	74	200	280	0	7.5	0.01	0.01	0	0	0.03
Per	60.1-60.9	8.2	73	190	260	0.2	7.4	0.01	0.01	0	0	0.03
Lev	59.5-61.8	8.1	71	190	260	0.1	7.2	0.01	0.01	0	0	0.03
Boz	57.2-57.5	8.8	34	130	130	0.1	0.6	0.01	0.01	0	0	0.03
LaD	59.8-60.2	6.7	23	220	1300	0	0.93	0.01	0.01	0.2	0	0.03
NM	56.4	6.6	27	120	600	0.8	4.3	0.01	0.01	0.7	0.1	0.03
BLV	57.0-57.3	6.6	27	120	600	0.9	4.7	0.01	0.01	0	0	0.03
WE	59.7-60.0	6.9	30	120	670	0.8	4.4	0.01	0.01	0.2	0	0.07
CW E	57.3-61.9	5.2	40	74	26	0.7	2.7	0.01	0.01	0.5	0	0.03
CW S	56.3-56.5	6.1	45	94	33	0.7	2.8	0.01	0.01	0.1	0.1	0.03
OS	54.0-64.0	8.2	110	300	26	1.4	3.3	0.01	0.01	0	0	0.03
TBV	59.3-60.7	8.4	110	290	21	2.1	3.3	0.01	0.01	0	0	0.03
Man	58.2-61.8	8.3	94	270	29	1	2.5	0.01	0.01	0	0	0.03
HP	56.6-58.3	9.2	120	410	44	2.1	4.3	0.01	0.01	0	0	0.08

[a] µM.
[b] mg

found only in New Zealand and Japan. The *Synechococcus* C1 lineage was dominant in Japan, and the *Synechococcus* C9 lineage was detected in all three countries studied. These unique geographic distributions indicated that thermophilic cyanobacteria are not evenly distributed globally, but have a history of geographic isolation.

Springs in the American Northwest included seven different geographical regions separated at the scale of springs within the same basin, to springs as distant as those in Oregon and Yellowstone National Park and the surrounding region (Figure 4.5 and Table 4.2). Hierarchical cluster analysis of hot spring chemical parameters showed that La Duke (LaD) and Mammoth Springs (Mam; consisting of New Mound Spring (NM), Bath Lake Vista (BLV) and White Elephant Back Spring (WE)) had significantly different chemistries compared to the other springs studied by Papke et al. (2003) (Figure 4.3), which was mostly due to higher concentrations of calcium, magnesium, carbonate and sulfate in these springs (Table 4.1A and B; gray shading). Springs in all other geographical regions in the Northwestern United States had water chemistries that clustered together, and formed an even larger cluster with most other springs located in New Zealand and Japan.

Analyses using the more rapidly evolving 16-23S rRNA ITS locus showed evidence of genotypic variation embedded within the B′-like 16S rRNA genotype (NACy10; Figure 4.3), which was specific to Yellowstone geyser basins, possibly indicating more localized geographic isolation (Figure 4.6). For instance, sequences from West Thumb (WT) and Lower Geyser Basin (LGB) (regions 6 and 5, respectively in

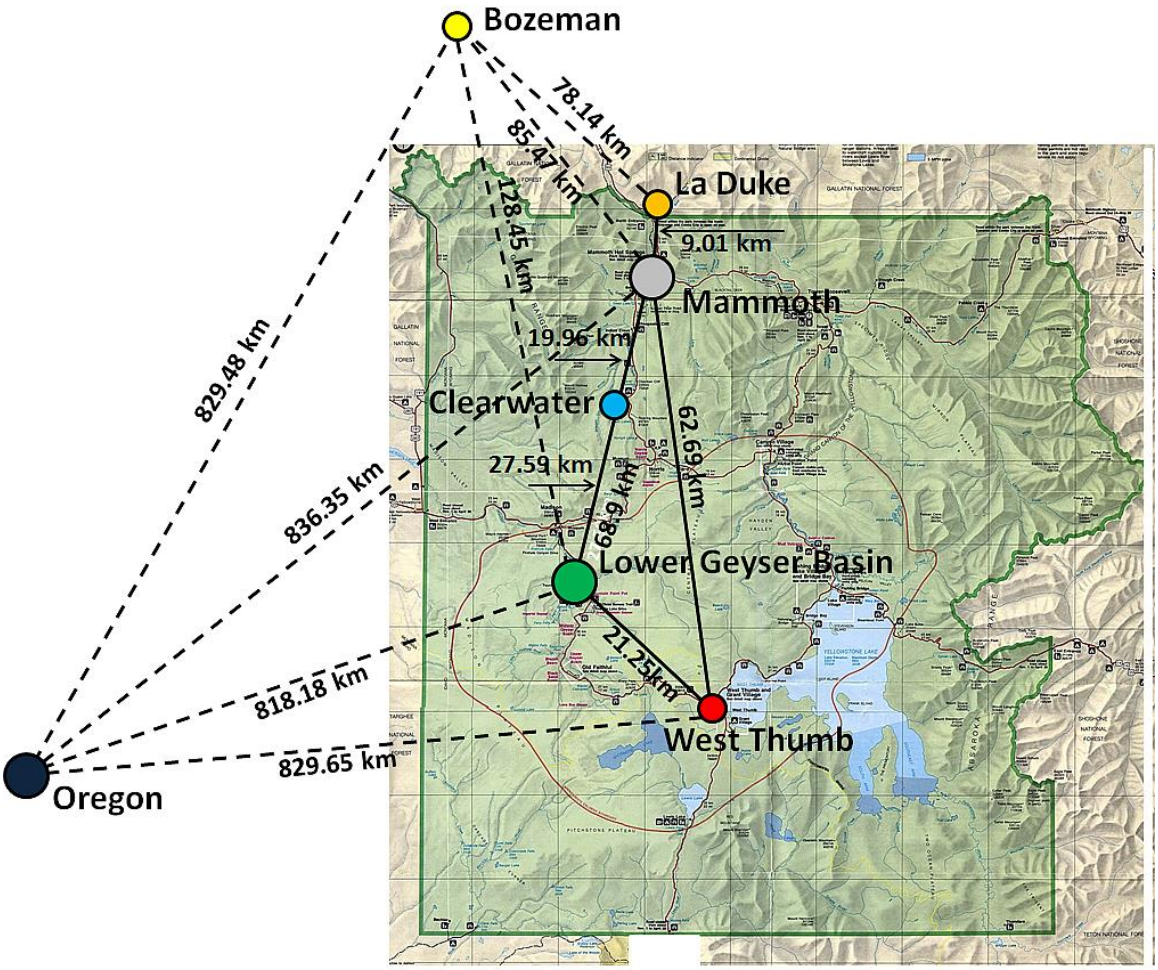


Figure 4.5. Approximate locations and geographic distances (km) among the 7 geographic regions studied across the Northwest United States by Papke et al. (2003).

Figure 4.4B) mostly formed clades separately from those obtained from the Mam, CW and LaD springs (regions 3 and 4). However, the low molecular resolution of the 16S-23S rRNA ITS locus was still possibly grouping many separate ecological species that can be detected with higher-resolution genetic markers like *psaA* (Chapters 2 and 3), obscuring biogeographic patterns across the Yellowstone region.

Understanding the impact that geographic isolation has on ecological species

Table 4.2. Summary of the samples and number of *psaA* sequences used for biogeographical analyses. Regions and springs shaded grey indicate different ecological settings that were used in ecological gradient analyses, and were not included in the biogeography study.

Geographic region	Latitude	Longitude	Spring	Average	pH value	Collection	Total	Total
Lower Geyser Basin	44° 32' 35''N	110° 46' 06''E	**Octopus Spring**	**54°C**	**8.2**	**30-May-96**	121	1253
Lower Geyser Basin	44° 32' 35''N	110° 46' 06''E	**Octopus Spring**	**60°C**	**8.2**	**30-May-96**		1535
Lower Geyser Basin	44° 32' 35''N	110° 46' 06''E	**Octopus Spring**	**64°C**	**8.2**	**30-May-96**		1479
Lower Geyser Basin	44° 32' 35''N	110° 46' 06''E	**Mushroom Spring**	**60°C**	**8.2**	**30-May-96**	90	2639
Lower Geyser Basin	44° 32' 35''N	110° 46' 06''E	**Twin Butte Vista**	**59.3-60.7°C**	**8.4**	**30-May-96**	99	2380
Clearwater	44° 47' 19''N	110° 44' 21''E	**Clearwater (east)**	**57.3-61.9°C**	**5.2**	**30-May-96**	82	5516
Clearwater	44° 47' 19''N	110° 44' 21''E	**Clearwater (south)**	**56.3-56.5°C**	**6.1**	**31-May-96**		1476
West Thumb	44° 24' 56''N	110° 34' 29''E	**Mantrap**	**56.4°C**	**8.8**	**31-May-96**	30	664
West Thumb	44° 24' 56''N	110° 34' 29''E	**Heart Pool**	**56.6-58.3°C**	**9.2**	**31-May-96**	28	2266
Mammoth	44° 57' 90''N	110° 42' 50''E	**White Elephant Back**	**59.7-60°C**	**6.9**	**31-May-96**	8	3270
Mammoth	44° 57' 90''N	110° 42' 50''E	**New Mound**	**56.4°C**	**6.6**	**31-May-96**	11	4384
Mammoth	44° 57' 90''N	110° 42' 50''E	**Bath Lake Vista**	**57-57.3°C**	**6.6**	**31-May-96**	13	4100
La Duke	45° 5' 25''N	111° 46' 28''E	**La Duke**	**59.8-60.2°C**	**6.7**	**31-May-96**	10	955
Bozeman	45° 39' 39''N	111° 11' 12''E	**Bozeman**	**57.2-57.5°C**	**8.8**	**8-Jul-96**	12	1125
Oregon	42° 11' 20''N	120° 20' 41''E	**Levee Spring**	**59.5-61.8°C**	**8.1**	**7-Jul-96**	17	2946
Oregon	42° 11' 20''N	120° 20' 41''E	**Perpetual Spring**	**60.1-60.9°C**	**8.2**	**7-Jul-96**	22	1418
Oregon	42° 11' 20''N	120° 20' 41''E	**Jack's Spring**	**50°C**	**8.8**	**7-Jul-96**	24	1297
Oregon	42° 11' 20''N	120° 20' 41''E	**Jack's Spring**	**60°C**	**8.1**	**7-Jul-96**		1625

[a] High-frequency sequences (≥50 100% identical copies in all combined samples).

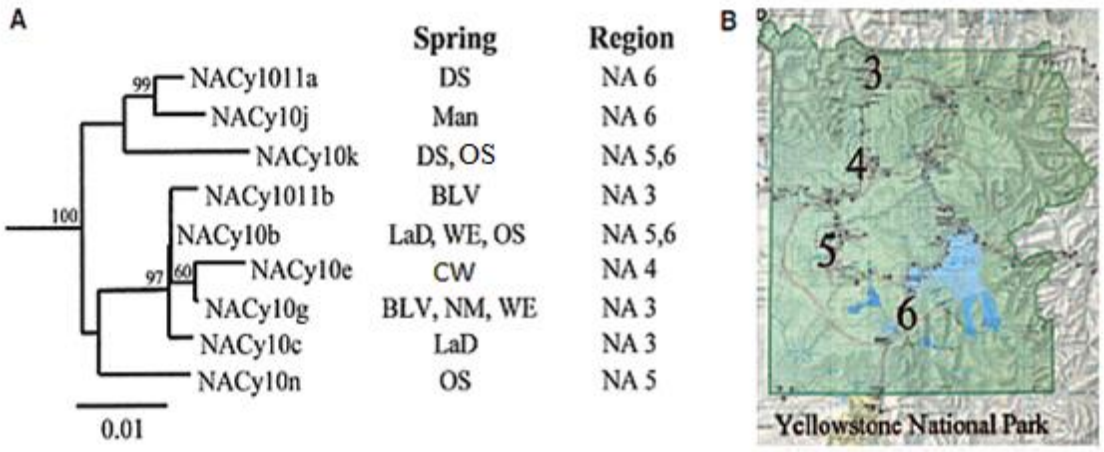


Figure 4.6. Phylogenies for internal transcribed spacer (ITS) variants detected in Yellowstone relative to springs and regions from which they were retrieved. (A) ITS genotypes with an identical 16S rRNA genotype (NACy10 corresponding to the type B′ 16S rRNA sequence from previous work (Ward et al., 1998)). (B) Map of Yellowstone National Park indicating regions sampled. Values at nodes indicate bootstrap percentages for 1000 replicates. Values less than 50% are not reported. Scale bar indicates 0.01 substitutions per site (modified from Papke et al., 2003).

populations within the regional species pool is important to understanding the forces that influence the ecological diversity and stability of these communities, and facilitates a better understanding of the *Synechococcus* cyanobacterial diversity contained within Yellowstone National Park, allowing for more directed measures in studying and protecting these natural resources.

To better understand the distributions of *Synechococcus* ecological variation in the American Northwest, I analyzed samples from Papke et al. (2003) using a morehighly resolving protein encoding locus (*psaA*). Geographic regions and springs were found to contain unique sequence and ecological variation, which decreased rapidly with geographic distance, indicating the importance of geographic isolation in the evolution of *Synechococcus* populations. Additionally, samples from ecologically distinct locations in

local environments were analyzed to determine how population distributions were affected by ecological, as opposed to geographical parameters.

Materials and Methods

Replicate mat samples for biogeographical analysis were collected from hot springs across the Northwest United States between 30 May and 8 July 1996 using a #4 cork borer (38.5 mm^2) by Papke et al. (2003), as summarized in Table 4.2. Samples used in biogeographical comparisons were collected at ~60°C and from near-neutral to alkaline pH sites in order to constrain ecological differentiation among populations, though some local ecological variation was present in the sample set. For instance, CW, Mam and LaD had lower pH than the other springs analyzed (5.2 to 6.9), but all other samples were from springs of pH 8.1 to 8.8. Also, as mentioned above, Mam and LaD had higher concentrations of calcium, magnesium, carbonate and sulfate compared to other Yellowstone springs. Data from chemical analyses are presented in Table 4.1, and samples were taken and analyzed as described in Papke et al. (2003). Additionally, in order to evaluate variation along ecological gradients, CW was sampled at 2 springs with different pH values, JS was sampled at 2 temperature-defined sites, and OS was sampled at 3 different temperatures along the effluent flow path (see Table 4.2). Samples were immediately frozen on dry ice (-20°C) in the field and kept frozen at -80°C until analysis in 2011.

DNA was extracted, and the *psaA* locus was PCR-amplified and sequenced using Ti454-barcoding technology, as described in Chapter 3 (Becraft et al., submitted).

Sequences were then trimmed to 302 base pairs to obtain the maximum number of sequences, cleaned and analyzed to identify high-frequency sequences (HFSs; ≥50 identical representatives across all combined samples) and associated low frequency sequences (LFSs; <50 copies in all samples) as described in Chapter 3 Materials and Methods. Phylogenies were constructed with MEGA 5.0 using the Neighbor-joining method (Tamura et al., 2011). Sequences were aligned using Clustal X (Larkin et al., 2007), and average pairwise nucleotide identities were calculated with MEGA 5.0. The percent contribution of each HFS (and associated LFSs) to all *psaA* sequences in each individual sample from ~60°C sample sites (ranging from ~57 to 62°C) was calculated (((HFSs+LFSs) / total sequences in sample)) x 100). Some samples (LaD, Boz, OS, Per and JS) could not be analyzed in duplicate due to failed sequencing reactions.

To analyze the distribution of B′-like HFS variation within ecological populations between chemically similar springs and regions, putative ecotypes (PEs; populations predicted through evolutionary simulation analysis to be ecologically distinct) were demarcated using a newer version of Ecotype Simulation (ES; an evolutionary simulation program based on the Stable Ecotype Model of species and speciation) currently being optimized by doctoral student Jason Wood and Dr. Fred Cohan and Dr. Danny Krizanc at Wesleyan University. Previously identified PEs B′5 and B′6 were demarcated together in this analysis. With the exception of the Ore-specific clade (which was analyzed separately), the A-like lineage was not analyzed by ES because reduced sequence length lumped many PEs described in Chapters 2 and 3. PEs contained dominant variant

sequences (DVs; identical sequences that made up the plurality of a PE clade), and PE DVs identified in Chapters 2 and 3 are indicated in Figures 4.7 and 4.8 for reference.

PEs were also demarcated with ES at sites sampled in order to analyze populations in regions or springs with multiple ecological settings (CW, JS and OS separately). Analysis of individual regions and springs with multiple ecological settings was conducted with the version of ES described in Chapter 3, Materials and Methods. PE percent population was calculated by dividing the total number of sequences within a PE detected in a sample (PE sequences = DV + all other HFSs + all LFSs) by the total number of sequences in that sample ((PE sequences in sample / all sequences in sample) x100). PEs mostly found in a particular geographical region are labeled according to that region. Additional PEs from MS and OS were demarcated in this analysis that were not identified in 2006 and 2008 studies and were labeled as 'newly demarcated' (ND) PE #, where # is numbered sequentially from 1 to n. Previously unidentified PEs that were mostly found in regions other than MS and OS were labeled with abbreviations corresponding to that region (e.g. Ore PE A1), and labeled sequentially.

Results and Discussion

To examine the biogeographical patterns of hot spring *Synechococcus psaA* sequence variants, a total of 43,757 (average of 1,492; standard deviation of 850 per sample) sequences were analyzed. These analyses yielded 133 HFSs that corresponded to samples from 13 springs within 7 different geographic basins (LGB, WT, CW, Mam,

LaD, Boz and Ore) that were separated by a minimum of 9 km, and a maximum of 836.35 km at the most distant sites (Figure 4.5).

As shown in Table 4.2, individual springs within LGB (OS, MS and BLVA) and CW had the most genetic diversity, with 121, 90, 99 and 82 HFSs, respectively. Man and HP (WT) had considerably less diversity, containing 30 and 28 HFSs, respectively. Lev, Per and JS (Ore) contained 17, 22 and 24 HFSs, respectively. WE, NM, BLV, LaD, and Boz (Mam and outlying springs) had the lowest genetic diversity observed, containing 8, 11, 13, 10, and 12 HFSs, respectively.

Distribution of HFS Variation in Springs of Similar Temperature and pH

HFSs were separated into A-like (Figure 4.7), and B′-like (Figure 4.8) *Synechococcus* phylogenies and the percentage of HFSs that were ≥1% of an individual sample are shown for each spring; HFSs that were <1% of the sample are indicated with asterisks. Individual springs are color-highlighted by geographic region (LGB, green; WT, red; CW, light blue; Mam, gray; LaD, orange; Boz, yellow; Ore, dark blue).

The LGB springs (MS, OS and TBV) had the most genetic variation spread across the A-like and B′-like phylogenies, and HFSs were mostly consistent with those identified at ~60°C in MS and OS in previous studies (Chapters 2 and 3). LGB springs contained 1 A-like and 4 B′-like clades that were specific to this region, but a large amount of the genetic variation detected in LGB springs was also detected in the neighboring WT and CW regions. In addition, WT region springs (HP and Man) also

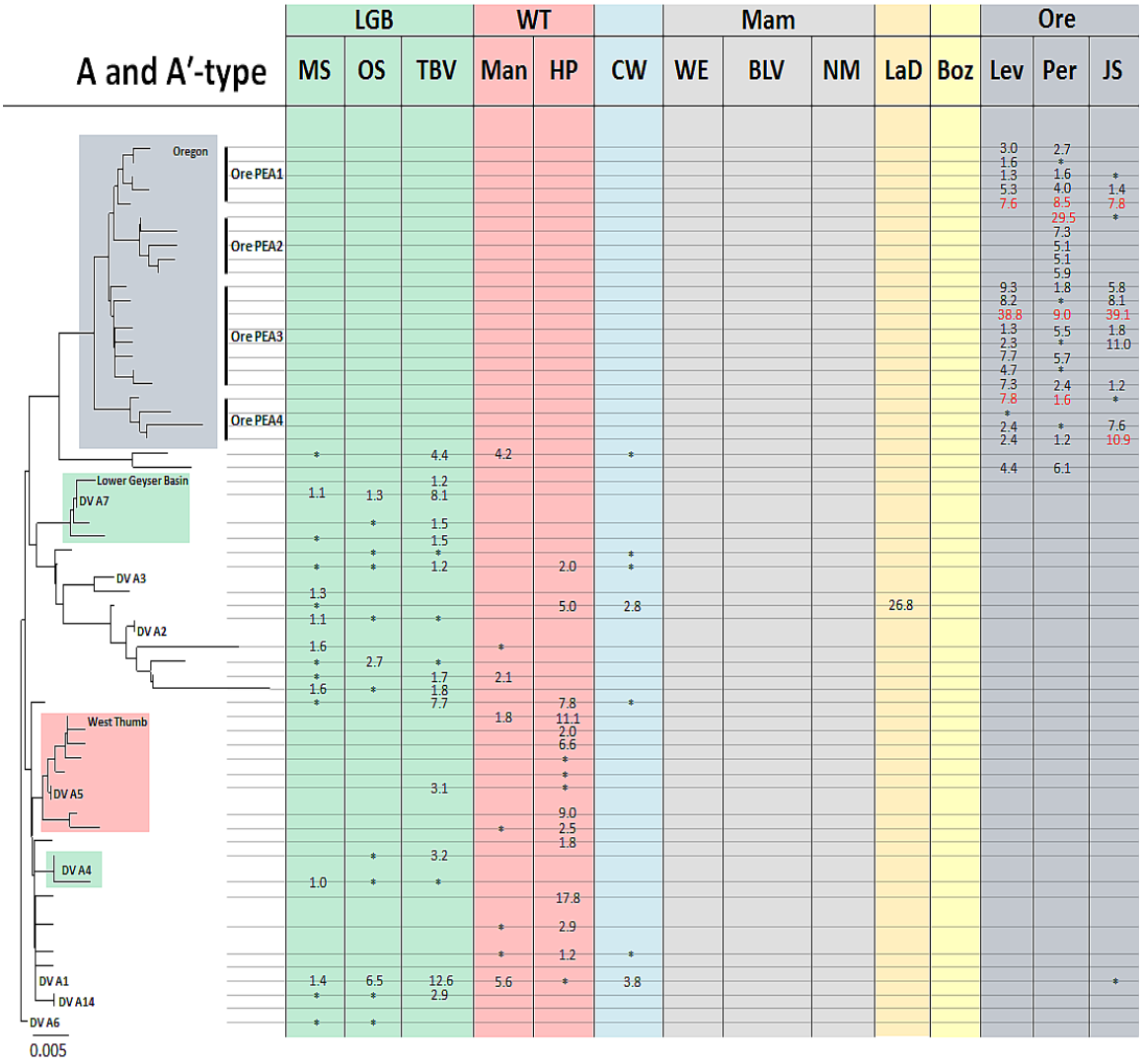
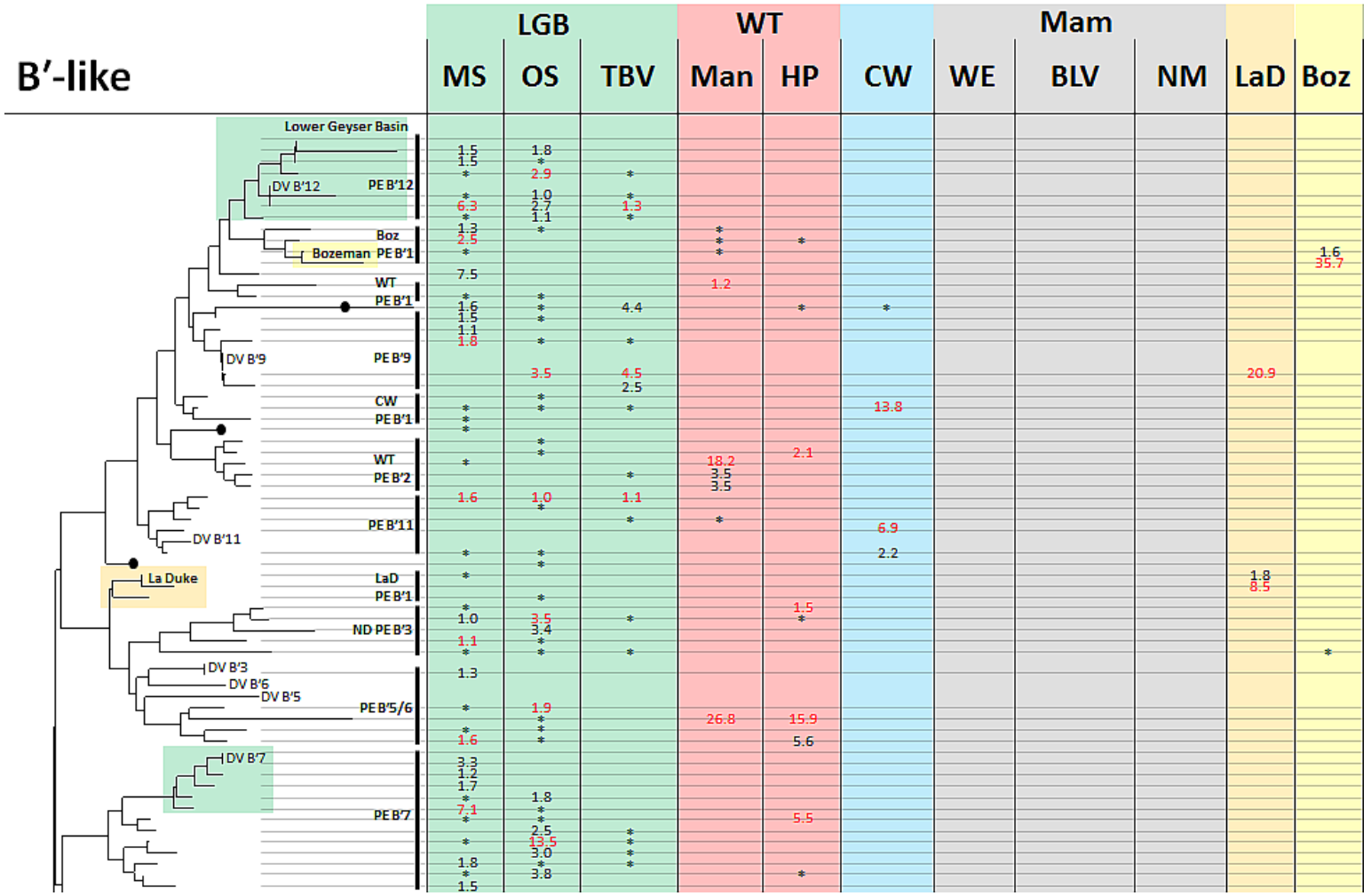


Figure 4.7. Phylogeny of A-like high-frequency sequences (HFSs) across 5 geographical regions in the American Northwest containing A-like diversity (Lower Geyser Basin (LGB), green; West Thumb (WT), red; Clearwater (CW), light blue; La Duke (LaD), orange; Oregon (Ore), dark blue). Percent of HFSs that are ≥1% of the sample are indicted adjacent to the sequence variants; asterisks indicate HFS percentages that were <1% of the sample. Color-shaded monophyletic clades indicate genetic diversity that is mostly specific to a region. Vertical bars represent putative ecotypes (PEs) demarcated by Ecotype Simulation; PE dominant variants (DVs) for each spring are colored red; DVs from PEs identified in Chapters 2 and 3 (vertical lines not shown for these PEs) are indicated in the phylogeny next to the identical HFS. Scale bar indicates .005 substitutions per site.


B′-like
LGB
WT
Mam
MS
OS
TBV
Man
HP
CW
WE
BLV
NM
LaD
Boz
Lower Geyser Basin
DV B′12
PE B′12
Boz
Bozeman
PE B′1
WT
PE B′1
DV B′9
PE B′9
CW
PE B′1
WT
PE B′2
PE B′11
DV B′11
La Duke
LaD
PE B′1
ND PE B′3
DV B′3
DV B′6
DV B′5
PE B′5/6
DV B′7
PE B′7

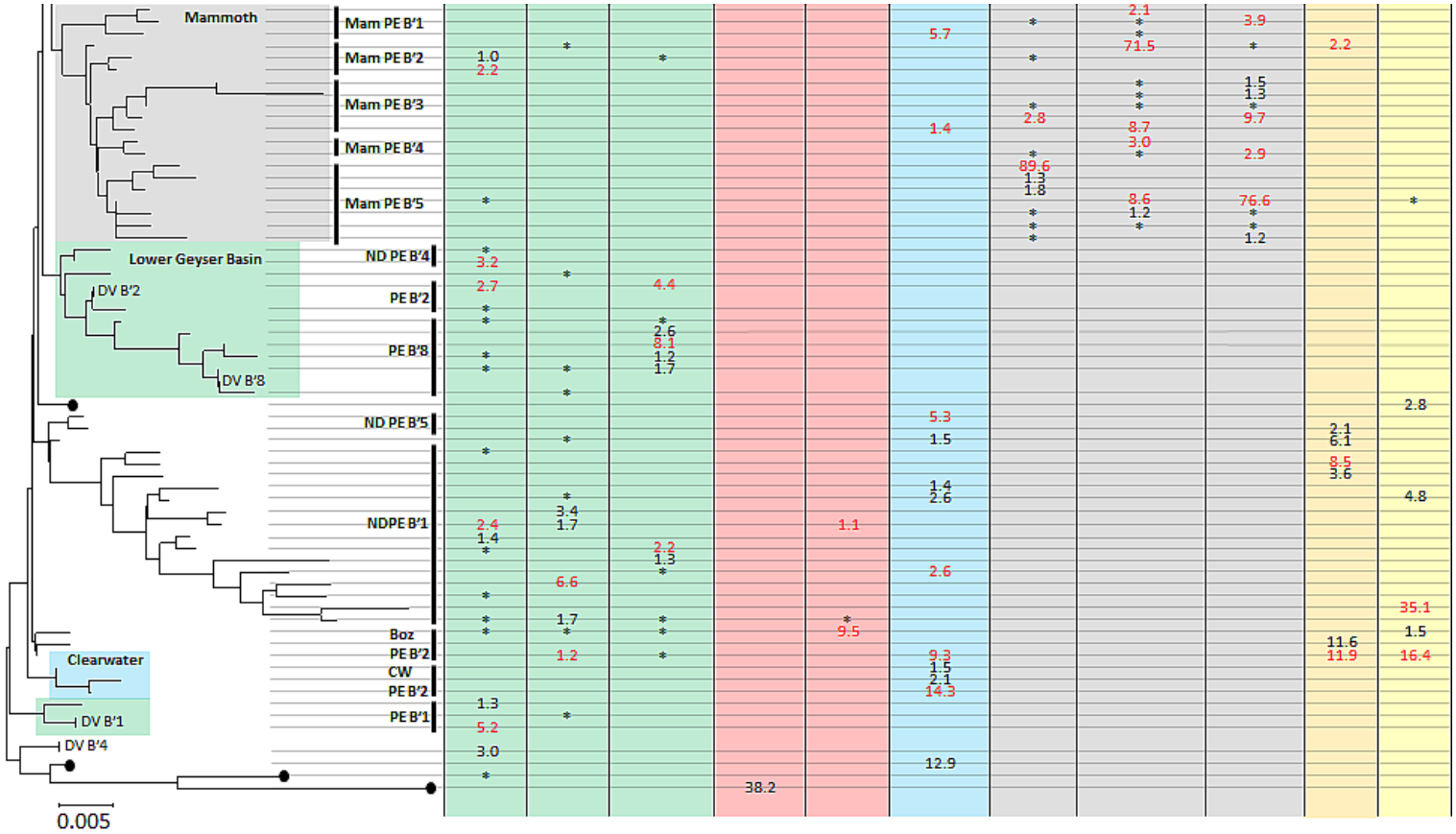


Figure 4.8. Phylogeny of B′-like high-frequency sequences (HFSs) across 6 geographical regions in the American Northwest containing B′-like diversity (Lower Geyser Basin (LGB), green; West Thumb (WT), red; Clearwater (CW), light blue; Mammoth (Mam), gray; La Duke (LaD), orange; Bozeman (Boz), yellow). Percent of HFSs that are ≥1% of the sample are indicated adjacent to the sequence variants; asterisks indicate HFS percentages that were <1% of the sample. Colored-shaded monophyletic clades indicate genetic diversity that is mostly specific to a region. Vertical bars represent putative ecotypes (PEs) demarcated by Ecotype Simulation; PE dominant variants (DVs) for each spring are colored red; DVs from PEs identified in Chapters 2 and 3 (vertical lines not shown for these PEs) are indicated in the phylogeny next to the identical HFS. Scale bar indicates .005 substitutions per site.

contained A-like and B′-like diversity, with some spring-specific HFSs. For instance, HP had 1 monophyletic clade of unique A-like HFSs that were mostly located at that spring (Figure 4.7, red box).

The most predominant HFSs from CW were specific to this spring, including one clade of relatively abundant HFSs in the B′-like lineage (CW PE B′2). However, some CW HFSs were also located in the neighboring LGB, Mam, Boz and LaD geographical regions in lower relative abundance. NM, WE and BLV only contained B′-like sequences that formed a clade of HFSs that were mostly unique to the Mam region (Figure 4.8; grey shading). LaD contained 11 genetically divergent HFSs, 4 of which were specific to this spring. Boz contained 9 genetically divergent B′- like HFSs, with the 2 most relatively abundant being spring-specific. Boz and LaD each contained 1 small monophyletic clade of relatively abundant HFSs that were unique to those regions (PEs Boz B′1 and LaD B′1). Other relatively abundant HFSs in Boz and LaD were also located in the neighboring Mam and LGB regions in lower relative abundance.

Sequences from hot springs located in Ore were monophyletic, genetically distinct and consisted of only A-like sequences. This was unexpected, since Papke et al. (2003) detected B′-like 16S rRNA sequences in JS (Figure 4.3) and Miller and Castenholz (2000) cultivated B′-like *Synechococcus* isolates from JS. It is possible that B′-like sequences are only located at average temperatures <60°C in JS. Another possibility is that genetic divergence within Ore populations was great enough to cause primer-pair mismatching, and the Ore B′-like sequences failed to PCR amplify.

HFS Relative Abundance vs. Geographic Distance

The average nucleotide divergence of HFSs within each geographic region ranged from 0.050 in LGB and CW, to 0.031 in WT, to 0.020 in Boz and LaD, to 0.019 in Mam and to 0.015 in Ore, and was correlated with the number of HFSs identified. The ANI was also calculated for HFSs between all regions, and no pattern of increased nucleotide divergence with increased geographic distance was observed (r^2 values of 0.03 for B′-like, and 0.4 for A-like sequences) (data not shown). However, there were clearly unique sequence variants at different springs across the Northwestern United States (Figures 4.7 and 4.8). With the exception of Ore HFSs, many sequences were equally genetically divergent from one another when the total variation was calculated, possibly due to the close geographical distance of the majority of the springs analyzed. This, in combination with the high level of HFS diversity in the LGB and CW regions, blurred the correlation between quantitative genetic differentiation and geographic distance.

All 13 springs analyzed contained unique HFSs, and many regions contained unique monophyletic clades of HFSs, though in many instances HFSs were located in multiple springs or regions. HFSs of high relative abundance were detected in neighboring springs and regions more often than geographically distant regions, and the relative abundances of HFSs declined with increased geographical distance. Figures 4.9 and 4.10 show the relative abundances of A-like and B′-like HFSs for each region analyzed, respectively. Geographic regions are separated along the horizontal axes (width), and geographic distance roughly increases with separation along the axis. Most

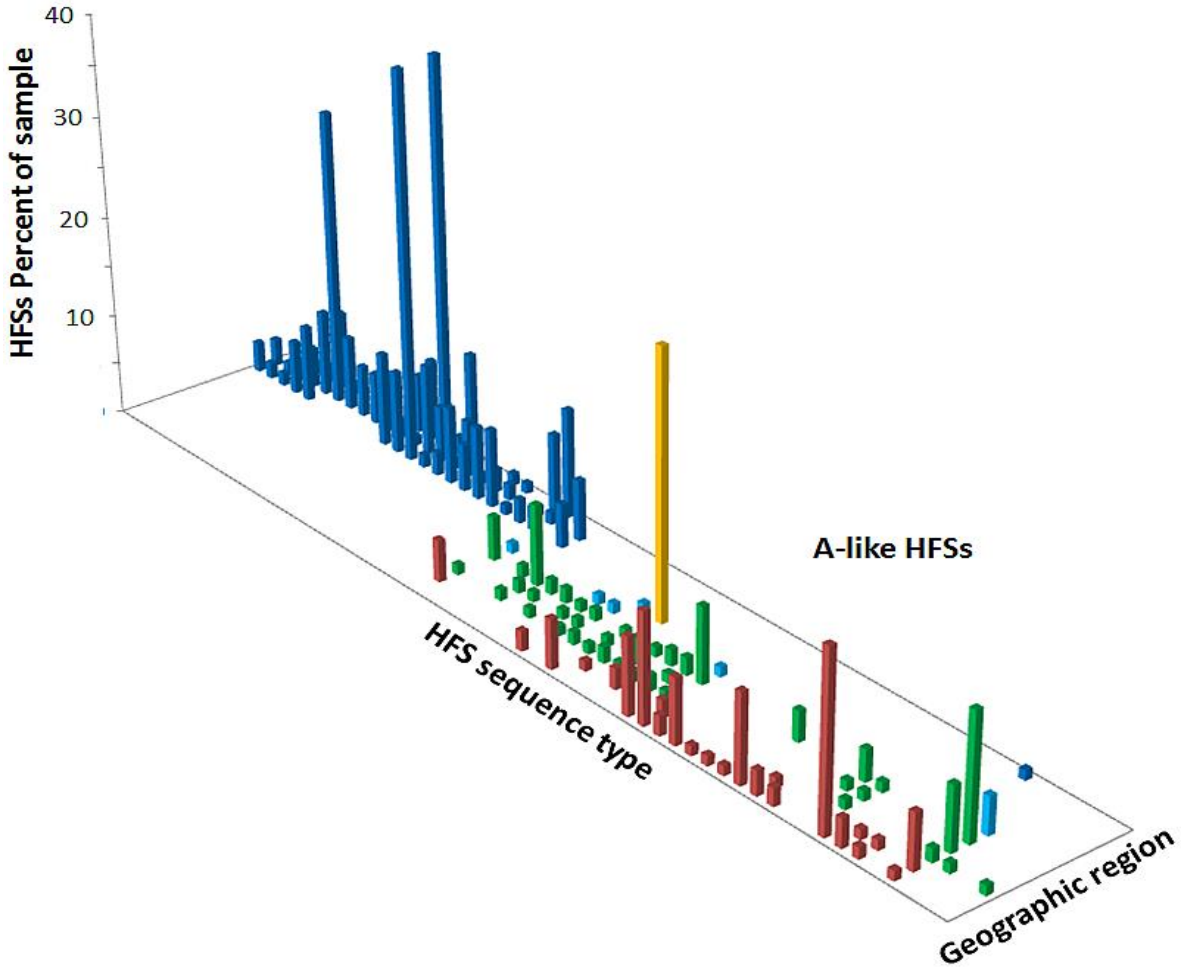


Figure 4.9. Relative abundances of A-like high-frequency sequences across 10 springs in the American Northwest (5 geographic regions) for each sequence type included in the analysis (West Thumb, red; Lower Geyser Basin, green; Clearwater, light blue; La Duke, orange; Oregon, dark blue). Distance roughly increases along the geographic region axis (width).

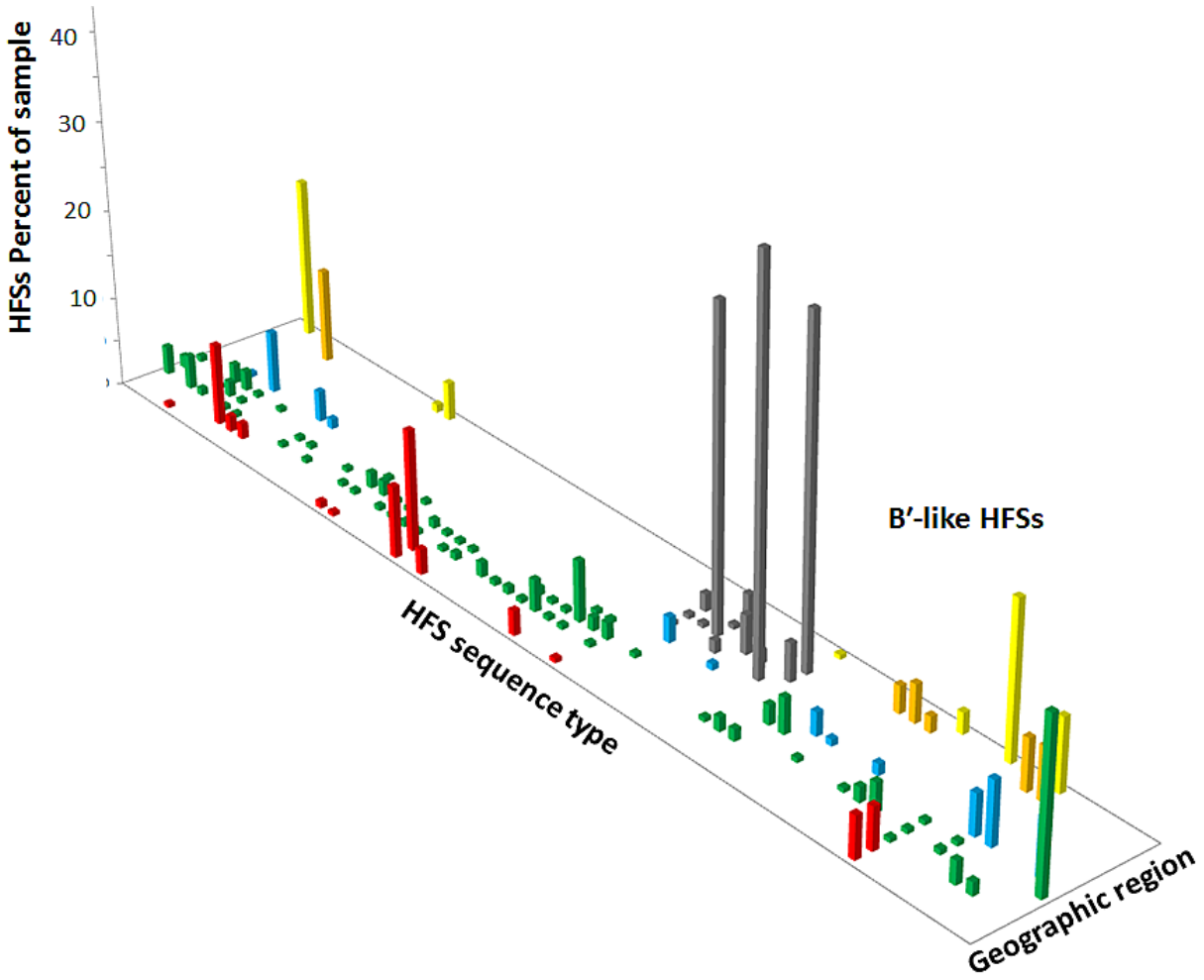


Figure 4.10. Relative abundances of B′-like high-frequency sequences across 11 springs in the American Northwest (6 geographic regions) for each sequence type included in the analysis (West Thumb, red; Lower Geyser Basin, green; Clearwater, light blue; Mammoth, gray; La Duke, orange; Bozeman, yellow). Distance roughly increases along the geographic region axis (width).

HFSs were only found in one spring or declined in relative abundance in adjacent springs and regions, further declining to zero quickly with increased geographic distance. Recurrent dispersal is insufficient to influence sequence frequencies directly. However, dispersal will be important in introducing new variation into a spring, either in populating a newly formed spring, or by introducing a sequence into an established spring where it previously did not exist.

Ecological Diversity Analyses

To analyze ecological populations within localized regions, HFSs from these regions were analyzed by ES and demarcated into PEs.

PE Diversity Among Chemically Similar Springs in the LGB and WT Regions.

B′-like HFSs were analyzed by ES to determine if different sequence variants within a PE are found in different basins. ES demarcated 32 B′-like PEs in the six geographic regions that contained B′-like diversity (Figure 4.8). ES calculated a periodic selection rate of 0.89 and an ecotype formation rate of 0.82 events/nucleotide substitution (also see Table 8.1, p.216). On average, despite the inclusion of individual regions and springs that had higher rates of periodic selection, ecotype formation events were calculated to be approximately as frequent as periodic selection events for B′-like *Synechococcus* populations in the American Northwest. The inclusion of geotypes in the analysis could have artificially increased the apparent ecotype formation rate calculated by ES.

Springs located in the LGB and WT regions have a high degree of chemical similarity (Figure 4.4, Table 4.1A and B), and provide the opportunity to observe the role

of geographic isolation on microbial populations. Different ecotypes among springs in the same region suggest that geographic isolation is more important than local ecological adaptation in these regions. B′-like PEs that are present in both the LGB and WT regions had different relative abundances and DVs (Figure 4.8; red highlighted numbers), and PEs were usually more abundant within an individual spring. The DV of PE B′11 was identical in all three LGB springs, but was different in CW. In contrast, the DV of ND PE B′1 was different in all three springs in LGB, and every region where it was located at >1% of the population. Sequence variants that were present in ≥2 springs were usually more dominant in one spring, and only 3 DVs were identical within a PE among all springs within a region (e.g. PEs B′2 and B′11 in LGB and PE B′5/6 in WT).

The presence of some and the absence of other PEs in both WT and LGB regions could indicate that different ecological species have different dispersal capabilities and ability to invade established communities. This could be due to species populations having different abundances along the vertical aspect of the mat, with surface-oriented populations having a greater chance of dispersal to neighboring regions (see Chapter 6, Figure 6.6, p.171). For example, PE B′7 is a surface population in EZ1 (Chapter 3, Figure 3.3, p.78), and was found in high relative abundance at OS and HP, and in lower relative abundance at MS and Man. This could be an example of a surface-oriented ecological population that was dispersed to MS and Man, followed by a change in the population structure over time due to periodic selection events, genetic drift and/or interspecies competition acting on the population separately in each spring (also see Chapter 5, Figures 5.5 and 5.6, pp. 140 and 144). Alternatively, there could be multiple ecotypes

contained within one PE, each better adapted to biotic and abiotic conditions in one or the other spring.

PE Diversity Among Springs in the Mam Region. ES demarcated five Mammoth Spring PEs (Figure 4.8; grey shading). Mam HFSs were also analyzed separately, and ES calculated a periodic selection rate of 3.5 and an ecotype formation rate of 0.9 events/nucleotide substitution (3.7:1), indicating that Mam populations have a higher periodic selection rate on average than other ecological populations identified in the American Northwest.

NM, WE and BLV each contained one predominant HFS that comprised the majority of the sequences in each spring (71.5 to 89.6%), a population structure that was unique to springs in the Mam region. ES demarcated 2 of the predominant HFSs (located in NM and WE) into Mam PE 5. If this demarcation is correct, this would suggest that populations of the same PE inhabiting different springs have different DVs, as was found in springs in the LGB. However, the DV at WE was located in a monophyletic subclade contained within Mam PE 5, and could be a young ecological population that was recently dispersed from NM to WE that was not demarcated separately by ES. Mam PE 2 contained as its DV the other predominant HFS, and was located mostly in BLV. Mam PEs 3 and 4 were lower in relative abundance and located primarily in both BLV and NM, though the PE DVs were different between the spring-specific populations.

The high relative abundance of the DV compared to the low relative abundance of other HFSs contained within these spring-specific populations could indicate that Mam PEs have recently experienced a bottleneck (e.g. recent dispersal and initial colonization).

Springs that form in the Mammoth region are typically ephemeral, forming and drying up over a period of a few years because of carbonate deposition. This could result in an environment where populations are initially dispersed to inhabit newly formed neighboring springs, take advantage of arriving at an uncolonized location first, and subsequently rise to abundance and diversify to a limited extent during the short-lifespan of the spring, without having to compete with locally adapted populations.

PE Diversity Among Springs in the Oregon Region. ES demarcated 4 Ore PEs (Figure 4.7, dark blue box). ES calculated a periodic selection rate of 1.73 and an ecotype formation rate of 0.97 events/nucleotide substitution, indicating that Ore populations have a lower periodic selection/ecotype formation ratio (1.7:1) than the ecological populations identified in Mam and MS (Chapters 3). Ore PE 2 was unique to Per, and contained the majority of the genetic diversity detected at that spring. The remaining genetic diversity contained in the Ore region-specific clade was dispersed across all 3 chemically similar springs (Figure 4.4), though the relative abundances of DVs and HFSs within PEs varied among springs. With the exception of Ore PE 4, PE DVs were the same in all 3 springs analyzed, indicating that frequent dispersal could be more important to ecological species distributions in this region. A single A-like sequence detected in Ore corresponded to a variant located primarily in the LGB, WT and CW regions (PE A1 DV in Figure 4.7), which could be the result of rare long distance dispersal or anthropogenic contamination.

PE and HFS Diversity in Clearwater Springs With Different pHs

CW was sampled at two springs (East and South) that had pH values of 5.2 and 6.1, and temperature ranges of 57.3 to 61.9°C, and 56.3 to 56.5°C, respectively (CWE and CWS, respectively) (Figure 4.11). ES predicted 21 PEs from the CW samples, 8 of which were based on a single HFS. ES calculated a periodic selection rate of 0.5 and an ecotype formation rate of 0.1 events/nucleotide substitution, yielding a periodic selection/ecotype formation event ratio of 5:1. CW PE 1 in Figure 4.11 consisted entirely of A-like sequences that were more relatively abundant at the high-pH site (CWS; 56.3 to 56.5°C). A-like diversity in CW contained a high number of relatively low-abundance HFSs, and only 10 of 48 HFSs comprised ≥1% of the sample. B′-like CW PEs had genetic variation found at both sample sites, though CW B′-like PEs 9, 10 and 12 were more relatively abundant at low pH (CWE), and B′-like CW PE 11 was more relatively abundant at high pH (CWS).

PE and HFS Diversity at Different Temperatures Within the Same Spring

JS was sampled at two temperature-defined sites (~50 and 60°C) along the effluent flow path (Figure 4.12A; also see Appendix C Figure C1). The low-temperature sample contained more HFS variation, which was also observed at CW and MS (see Chapters 2 and 3). ES predicted 6 PEs at JS, 2 of which were based on a single HFS. ES calculated a periodic selection rate of 0.6 and an ecotype formation rate of 0.4 events/nucleotide substitution, indicating that ecotype formation rates are almost as

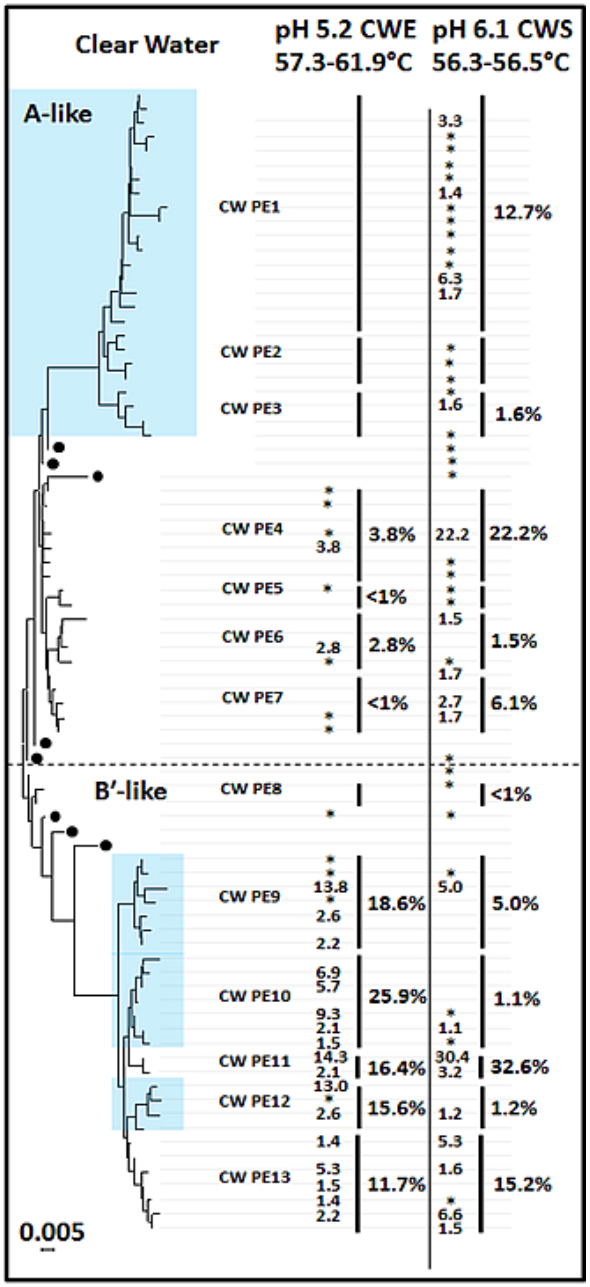


Figure 4.11. Phylogeny based on high-frequency sequences (HFSs) comprising ≥1% of the samples at Clearwater Spring (CW). CW samples corresponded to pH 5.2 and a temperature range of 57.3 to 61.9°C (CWE, left), and pH 6.1 and a temperature range of 56.3 to 56.5°C (CWS, right). A-like HFSs are located above, and B′-like HFSs are below the dotted horizontal line. Vertical bars indicate putative ecotypes (PEs) demarcated by Ecotype Simulation; black dots represent PEs based on single HFSs. Percent HFSs in the total sample is indicated to the left, and PE percent population is indicated to the right of the vertical bars. Shaded boxes indicate HFS clades that are mostly or completely site-specific. Scale bars indicate 0.005 nucleotide substitutions per site.

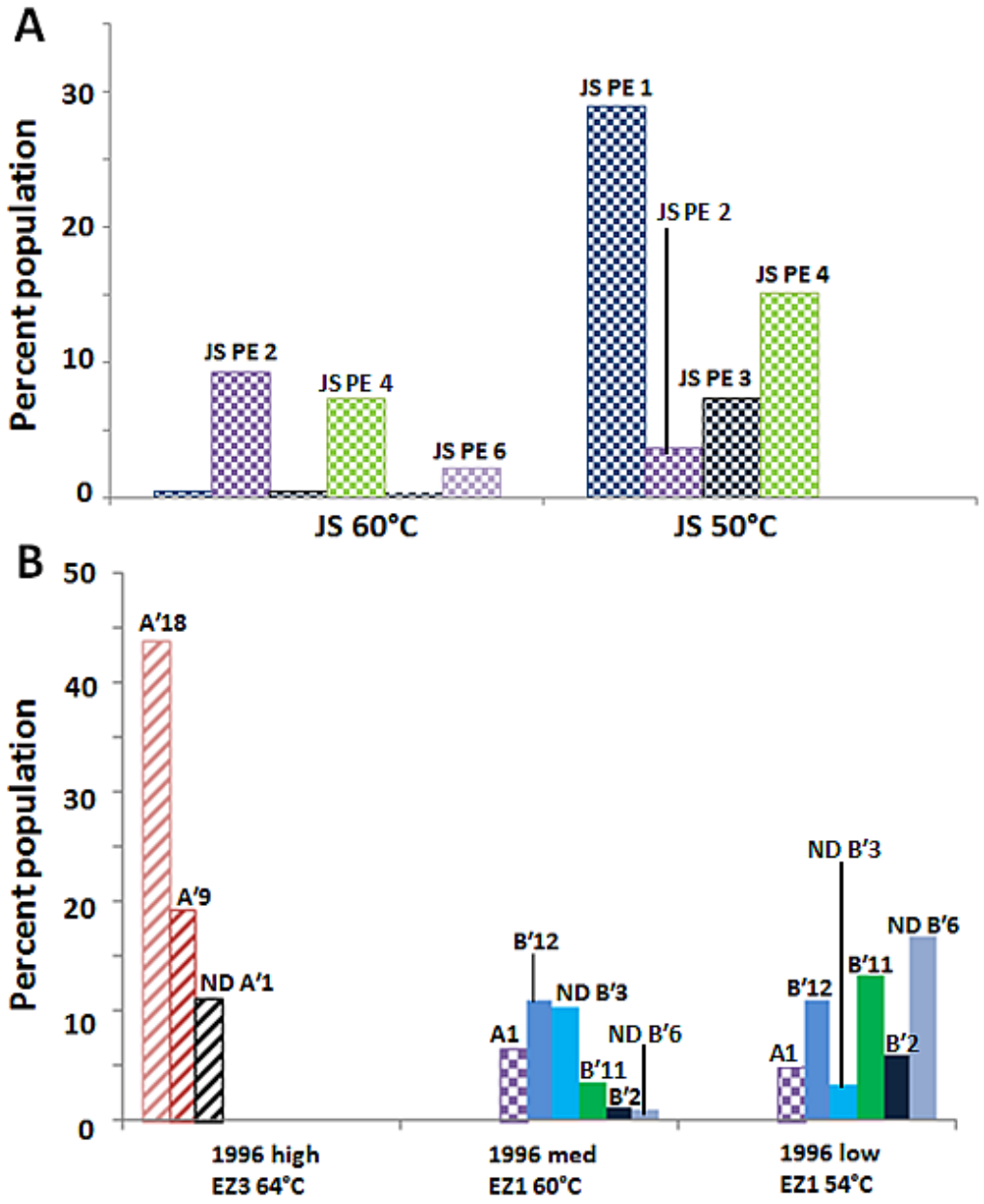


Figure 4.12. Percent population of abundant putative ecotypes (PEs) at (A) two temperature defined sites along the Jack's Spring effluent flow channel and (B) three temperature-defined sites along the Octopus Spring effluent flow channel. Newly demarcated PEs are labeled ND; colors correspond across temperature sites. Striped bars represent A′-like PEs, dotted bars A-like PEs, and solid bars B′-like PEs.

frequent as periodic selection rates in JS, and similar to ES predictions for the entire Ore region. JS PEs 1 and 3 were clades consisting of HFSs located almost exclusively at low-temperature site, while JS PE 6 was exclusively located at the high-temperature site. JS PE 2 was more relatively abundant at the high-temperature site, though HFSs within JS PE 2 were also found downstream in lower relative abundances. JS PEs 2 and 4 could be ecological populations that can persist at both temperature sites, though are better adapted to one site or the other, or something in between.

OS was sampled at 3 locations along the effluent flow path, including high- (~64°C), medium- (~60°C) and low-temperature (~54°C) sites (Figure 4.12B). As observed with MS flow path distributions (Chapters 2 and 3; Figures 2.3 and 3.2, pp. 47 and 77), there appeared to be ecotone boundaries where genetic variation changes along the effluent flow path, with A′ PEs located at ≥64°C (EZ3), and B′-like and A-like (in lesser abundance) PE diversity located at 60°C and 54°C (EZ1). In addition to PE A′9, ND PE A′1 and PE A′18 were more relatively abundant in OS EZ3, and PE A′18 was nearly 50 percent of the sample's total diversity. ND PE B′3, ND PE B′6 and PEs B′11, B′12 and A1 were the most relatively abundant PEs in OS EZ1 (also see Chapter 5, Figure 5.6, 1996, p.144). With the exception of PE B′12, PE relative abundances were different between the 60°C and 54°C EZ1 sites.

Conclusions

Advances in molecular biology are providing a multitude of evidence that the once widely held belief that "everything is everywhere, and the environment selects" is

not correct for all taxa. Here I present evidence that hot spring *Synechococcus* populations have strong barriers to frequent geographic dispersal. High-throughput sequencing of *psaA* variation across the Northwestern United States better resolved biogeographical patterns of *Synechococcus* sequence variants first observed by Papke et al. (2003) and enabled analysis of the population genetic structures of ecological species.

Many regions contained monophyletic clades of unique HFS variation. HFSs that were shared between geographical regions, and between springs within the same region, varied greatly in their relative abundances. There were higher levels of migration of HFSs between neighboring springs and geographic regions, and the relative abundances of HFSs decreased with geographic distance. This could be because HFSs corresponding to ecological populations can migrate short distances and persist at chemically similar springs (e.g. B′-like populations between LGB and WT), but do not reach abundance because they are not as well adapted as are the spring-specific biota to biotic and abiotic conditions of the local community (i.e. are unable to invade due to interspecific competition). All springs contained unique HFSs, even if only slightly different from other variants detected in nearby springs. The Oregon region contained unique A-like variation, and CW, LaD, LGB and WT all contained closely related A-like HFSs that varied in presence and relative abundance across springs and regions (e.g. LGB compared to WT). B′-like PEs were completely or mostly region specific (e.g PE B′7), and PEs had different HFSs and genetic structures among springs and regions (e.g. PE B′5/6 and WT PE B′2). This could indicate that the forces of periodic selection and genetic drift are acting upon ecological populations independently in these local communities (hot

springs), and despite short geographic separation, these populations are changing separately and irreversibly diverging from one another due to geographic isolation. As an exception, Ore PEs 1 and 3 had the same DV among the 3 springs analyzed, indicating that frequent dispersal could have more impact on the ecological species in this region, or that these springs are closer in proximity.

Some PEs in the chemically different Mam region contained one highly abundant HFS with minimal surrounding variation, possibly because these populations have gone through a recent bottleneck event (e.g. recent dispersal to a newly formed spring) since carbonate-depositing springs are ephemeral in nature. The high relative abundance of one sequence variant in relation to others within a population indicates that these are possibly young ecological/geographical species (also see Chapter 8).

Ore and CW both contained ecological species that were mostly region-specific, though many populations shared variation between springs in varying relative abundances. The relative abundances of populations changed along environmental gradients (temperature) at JS and OS, similar to patterns observed at MS (Chapters 2 and 3). OS contained many of the PEs identified in MS, although the relative abundances varied greatly between springs (e.g. PEs A′9 in MS compared to A′18 in OS EZ3, and B′9 in MS compared to B′12 in OS EZ1) (also see Chapter 3, Figure 3.2, p.77). While some PE DVs were identical across all springs within a region (e.g. PE B′11), DVs were most often different between springs, and no DV that was present in ≥2 regions was identical.

The ratio of periodic selection to ecotype formation events varied somewhat between springs and geographical regions. The ratio of periodic selection/ecotype formation events was 1.78:1 in Ore (1:1.5 at JS), 2.9:1 in MS, 3.7:1 in Mam, and 5:1 in CW. Assuming that mutation and recombination rates are constant, these ratios might indicate that populations are more 'stable' (i.e. more time has lapsed since dispersal or the last periodic selection event) in Ore. The amount of genetic diversity within ecotypes should be dependent upon the amount of time that has passed since the last bottleneck event (e.g dispersal to newly formed spring or periodic selection event), and older populations should contain more sequence variants surrounding a less-abundant dominant variant (Chapter 8, Figure 8.1, p.211).

The B′ *Synechococcus* species diversity in the American Northwest yielded a periodic selection/ecotype formation ratio of ~1:1, indicating that ecotype formation events are nearly as frequent as periodic selection events in these populations. Higher rates of ecotype formation events were also observed for *Bacillus subtilis* populations by Koeppel et al. (2013). Alternatively, the higher rate of ecotype formation could be due to the inclusion of multiple geotypes (i.e. ecologically interchangeable species separated by geographic distance) that have diverged apart, but maintain the same ecology in chemically similar springs.

CHAPTER 5

SEASONAL AND YEARLY CHANGES OF *SYNECHOCOCCUS* ECOLOGICAL POPULATIONS IN MUSHROOM SPRING AND OCTOPUS SPRING MAT COMMUNITIES

Introduction

Population dynamics is the long-term study of changing abundance and composition of species within communities due to ecological processes, such as environmental change or dispersal from neighboring communities in the regional species pool (Morin, 2011; Chapter 4 Figure 4.1, p.94). Population genetics is the study of gene frequencies and change within populations over time (Lincoln et al., 1998) due to evolutionary processes. For instance, periodic selection events might quash diversity within ecological populations, mutation and/or recombination might introduce new variation into populations, and drift might cause genetic variants to rise and fall in abundance. Furthermore, populations can go extinct and speciation might give rise to new ecologically specialized populations. All of these forces might act upon microbial populations and local communities, influencing the diversity and stability of the ecological species they contain.

Microbial populations have been shown to change seasonally in relation to fluctuating environmental conditions (DuRand et al., 2001; Thompson et al., 2004; Giovannoni and Vergin, 2012). Although Mushroom Spring (MS) and Octopus Spring (OS) have been extensively studied for over 30 years, these studies have mostly treated

the systems as if they were static. To determine if populations in these mats changed in response to changing environmental conditions, Ferris and Ward (1997) analyzed the population diversity at OS seasonally using denaturing gradient gel electrophoresis of partial 16S rRNA gene segments. Results for June and December are shown in Figure 5.1. Light availability changes over the diel cycle, and total available light at OS changes seasonally. The day length is considerably shorter in winter months, and the 24 h average irradiance is about 25% of that in the summer (Figure 5.2). Ferris and Ward (1997) found that despite differences in light intensity *Synechococcus* populations did not change seasonally. However, the study was complicated by the fact that the 16S rRNA molecule lacks the molecular resolution needed to accurately demarcate ecological species (Melendrez et al. 2011; Becraft et al., 2011, submitted; Chapter 3). In addition, disturbance events are more stochastic environmental changes that might affect species composition and structure in time (Brock and Brock, 1968; Ferris et al., 1997).

Using Ti454-barcoding and sequencing of *psaA* sequences, I was able to better understand the temporal variation of *Synechococcus* populations in MS and OS at greater molecular resolution than is provided by 16S rRNA sequences. I analyzed samples collected in different seasons, and reexamined samples collected over a 20 year period by various members of the Ward Lab in order to more clearly resolve the population dynamics of ecological species in response to fluctuating environmental conditions or environmental perturbations. Ti454-barcoding and sequencing also provided the ability to observe individuals within ecological populations, and thus to examine the population

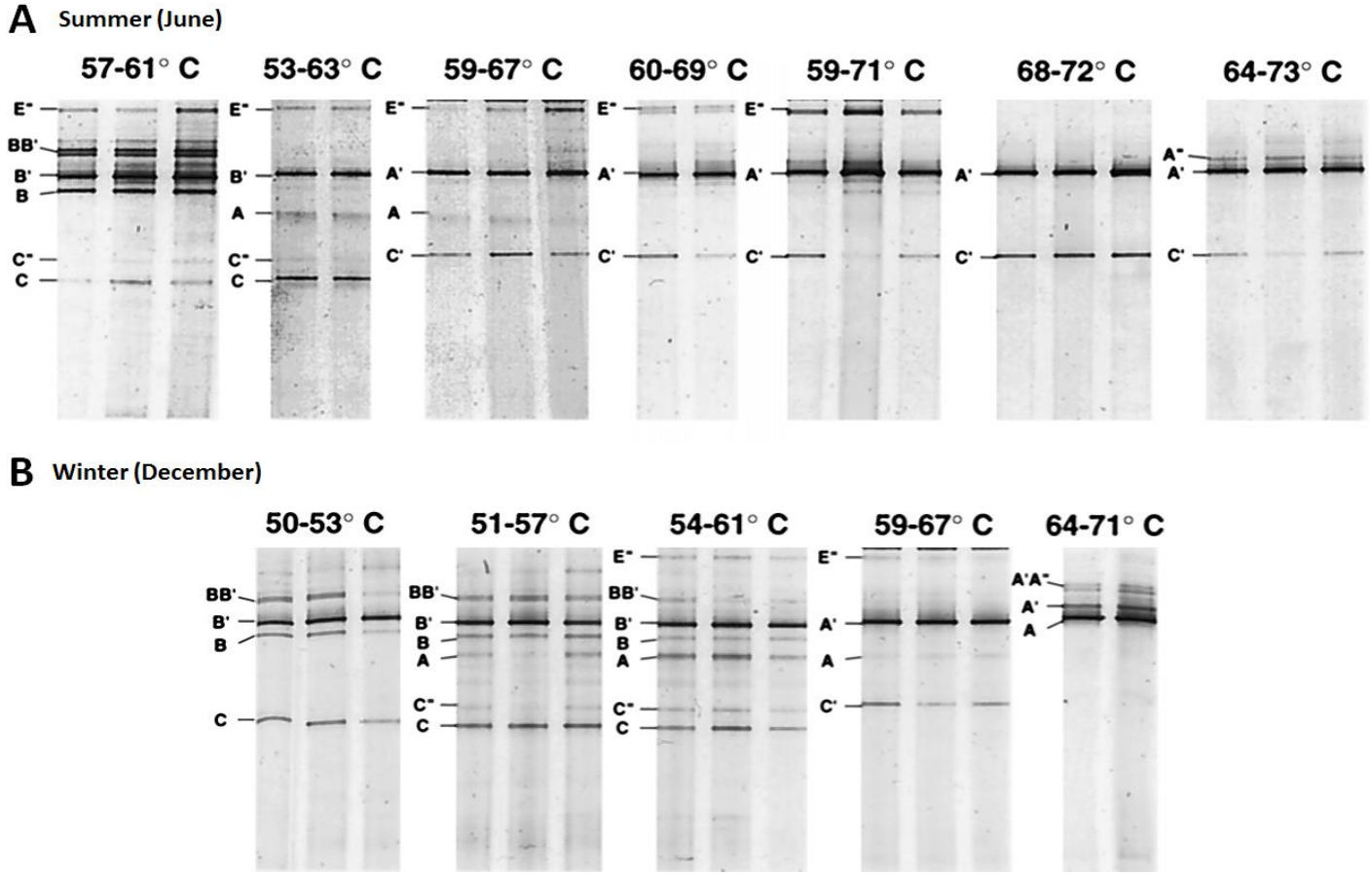


Figure 5.1. Denaturing gel gradient electrophoresis profiles of PCR-amplified 16S rRNA gene segments from DNA extracted from Octopus Spring microbial samples at temperature-defined sites along the effluent channel in (A) June and (B) December. Single letters correspond to sequence types; genotypes of interest to this dissertation are A, A′ and B′. (modified from Ferris et al., 1997).

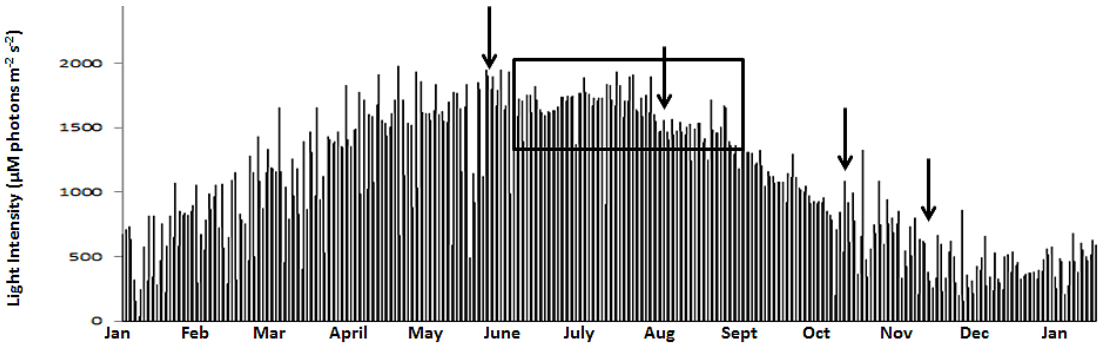


Figure 5.2. Light intensity averaged over 24 h measured from January 12 2011 to January 12 2012 at Mushroom Spring (MS). Inset box identifies light intensities over a series of samples collected between June 1996 to September 2009 at MS in ecological zone (EZ) 1 (see below). Arrows indicate months in which Octopus Spring samples were collected in EZ1.

genetics of ecological species, and the forces that influence within-species variation over time within these local communities.

Materials and Methods

Sampling

Samples for comparison of population structure in summer and winter were collected by collaborating doctoral student Shane Nowack on 2 June 2011, and 15 December 2011, as a part of a collaboration between the Ward Lab and the laboratory of Isaac Klapper (MSU Department of Mathematics and Temple University Dept. of Mathematics). A #2 cork borer (19.6 mm^2) was used to collect mat samples at 4 spatially fixed sites along the effluent flow path at MS and OS. Temperatures at sample collection sites were recorded every 5 minutes, and averaged over 30 minutes, using the DS1922T Temperature Logger iButton with 8KB memory (Maxim Integrated, San Jose, CA) and are reported in Table 5.1.

Samples from different years were collected between 1989 and 2009 by members of the Ward Lab, using a #4 cork borer (38.5 mm^2) at temperatures ranging from 55 to 64°C and are summarized in Table 5.2. All samples were immediately frozen on dry ice (-20°C) in the field and kept frozen at -80°C until analysis in 2011.

Molecular Methods

DNA was extracted as described in Chapters 2 and 3 (Becraft et al., 2011; Becraft et al., submitted), and the *psaA* locus was amplified and sequenced with the 'centerfowardprimer' (see Chapter 3 Material and Methods) using Ti454-barcoding technology. Yearly mat samples were sequenced at the J. Craig Venter Institute as described in Chapter 3 Materials and Methods, and seasonal mat and water samples were sequenced at the Research and Testing Laboratory, Lubbock, TX (Rhoads et al., 2012).

Sequences were trimmed to 302 base pairs to obtain the maximum number of sequences, cleaned and analyzed to identify high-frequency sequences (HFSs; ≥50 identical representatives across all combined samples) and associated low-frequency sequences (LFSs; <50 copies in all samples) as described in Chapter 3 Materials and Methods. Phylogenies were created in MEGA 5.0 using the neighbor-joining method as described in Chapter 4 Materials and Methods.

Ecotype Demarcation

Putative ecotypes (PEs; populations predicted through evolutionary simulation analysis to be ecologically distinct species) were demarcated using Ecotype Simulation

Table 5.1. Collection sites and temperatures for seasonal study. Average, minimum and maximum temperatures taken over a 24 hour period at Mushroom Spring. Sample sites progress downstream along the effluent flow path from site 0 (just downstream from the source) to site 4. Average ambient temperature is also reported.

	Site 0	Site 1	Site 2	Site 3	Site 4	Ambient °C
June 6 2011						
24 hr-avg	¥	65.40°C	62.27°C	57.38°C	¥	10.26°C
Max		66.83°C	63.99°C	60.47°C		25.66°C
Min		62.68°C	59.54°C	52.91°C		2.65°C
°C	64.1°C				42.1°C	
December 15						
24 hr-avg	¥	¥	62.36°C	58.30°C	51.03°C	-5.40°C
Max			62.53°C	58.77°C	51.66°C	-1.47°C
Min			61.79°C	57.41°C	49.49°C	-7.58°C
°C	66.3°C	66°C				

¥ Missing data are due to failed temperature data-loggers, usually because of prolonged exposure to high temperatures, and point measurements taken at the time of collection are reported in these cases.

Table 5.2. Summary of samples and sequences used for temporal analyses in Mushroom Spring and Octopus Spring from 1989 to 2009.

Spring	Average temperature	Collection date	Total HFSs [a]	Total sequences
Octopus Spring	**55°C**	**16-Oct-89**	88	1498
Octopus Spring	**60°C**	**16-Oct-89**	131	1650
Octopus Spring	**59-61°C**	**8-Aug-91**	127	1396
Octopus Spring	**54°C**	**30-May-96**	121	1253
Octopus Spring	**60°C**	**30-May-96**	90	1535
Octopus Spring	**54°C**	**18-Nov-99**	132	1154
Octopus Spring	**52-64°C**	**18-Nov-99**	79	2296
Octopus Spring	**55-64°C**	**23-Jun-05**	66	2459
Mushroom Spring	**56.4°C**	**12-Jun-96**	114	1157
Mushroom Spring	**54-60°C**	**7-Oct-99**	89	1446
Mushroom Spring	**56.4°C**	**11-Aug-06**	137	1592
Mushroom Spring	**60°C**	**13-Dec-07**	79	1504
Mushroom Spring	**60°C**	**9-Sep-08**	120	1636
Mushroom Spring	**60°C**	**18-Jun-09**	229	12492

[a] High-frequency sequences (≥50 100% identical copies in all combined samples).

(ES; an evolutionary simulation program based on the Stable Ecotype Model of species and speciation; see Chapter 1), and PE population percentages were calculated as

described in Chapter 3 Materials and Methods. A-like PEs were demarcated using the fine-scale approach, and A′-like and B′-like PEs were demarcated using the conservative approach, as described in Chapter 3 Materials and Methods (also see Appendix B, Section 1). Most PEs contained a dominant variant (DV; identical sequences that made up the plurality of a PE clade) surrounded by less-abundant HFS variation and associated LFSs. Only HFSs that were ≥1% in any sample were included in phylogenies. Total PE percentages sum to between 70 and 90% for each sample, and the remaining diversity consisted of sequences that were <1% of the total in any sample analyzed over the 20 year period (data not shown).

Results and Discussion

Seasonal Effects on PE Distributions in Mushroom Spring Mat Samples

Samples from December and June were compared to determine whether PE population distributions and relative abundances differed (Figure 5.3). Water temperature was nearly identical between June and December, despite differences in ambient temperature (Table 5.1). In June (Figure 5.3, left), the upstream mat samples (sites 1 and 2) were indicative of EZ3 and contained mostly A′-like PEs, and PEs A′8 and A′9 were the most relatively abundant populations. The lower temperature samples (sites 3 and 4) were indicative of EZ1, and PEs A1 (site 3), B′2, B′3, B′7, B′8 and B′9 were the most relatively abundant populations. Overall, there was a general downstream progression from A′-like, to A-like to B′-like PEs.

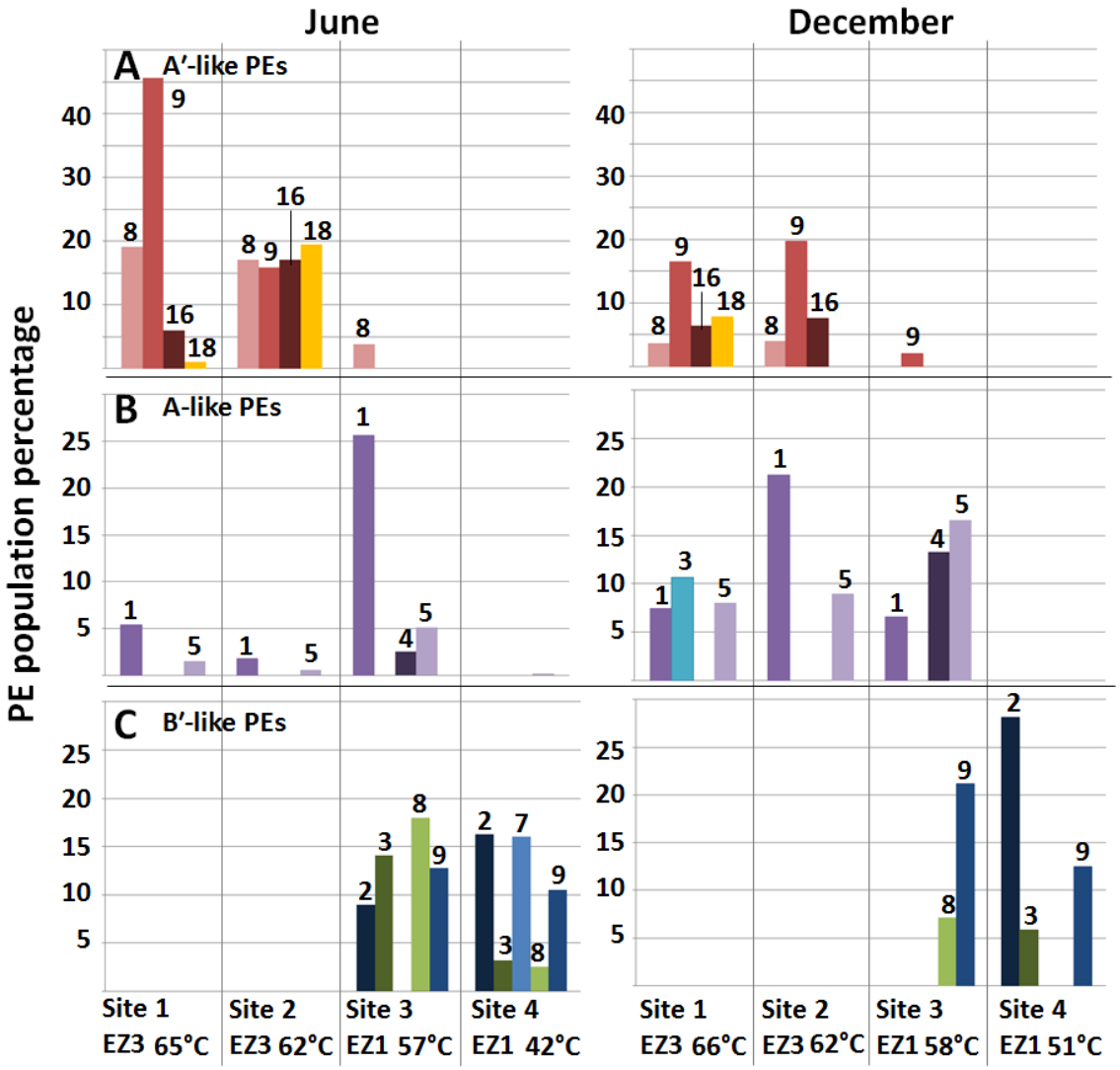


Figure 5.3. Percent population of relatively abundant (A) A′-like (B) A-like and (C) B′-like *Synechococcus* putative ecotypes (PEs) in mat samples along the effluent flow path in Mushroom Spring in June (left) and December (right). Ecological zones (EZs) were determined by comparison of summer samples to flow path samples collected in 2006 and 2008 studies (Chapters 2 and 3). Temperatures are given for each sample site (also see Table 5.1).

In the December samples, PEs A′8 and A′9 were lower in relative abundance at site 1, and PEs A3 and A5 were higher in relative abundance (Figure 5.3, right). At site 2 A′-like PEs A′8 and A′18 were lower in relative abundance, and A-like PEs A1 and A5

were higher in relative abundance. At site 3 PE A1, which was the only predominant A-like PE found in EZ1 in summer, was lower in relative abundance (Chapter 3; Figure 3.2, p.77), and A-like PEs A4 and A5 were higher in relative abundance. Additionally, PEs B′2, B′3 and B′8 were lower in relative abundance at site 3. At site 4 PE B′7 was <1% of the total sample, despite comprising >15% of the *Synechococcus* population in June, and PE B′2 was higher in relative abundance. A-like PEs were in higher relative abundance in winter at both warmer and cooler temperatures (EZ3 and EZ1, respectively), and appeared to cross ecotone boundaries and utilize positional locations occupied by A′-like and B′-like PEs in the summer.

Despite similar water temperatures, PE relative abundances were different in December and June, which could indicate that temperature only partially determines ecotype distributions along the effluent flow path. While there are currently no vertical distributions analyses for winter mat samples, at lower light intensities deeper PE populations (e.g PE A4; Chapter 3, Figure 3.3, p.78) might increase in relative abundance higher towards the mat surface where they are better adapted to the reduced irradiance conditions, and have a higher relative abundance in the whole-mat sample. Surface mat conditions in the winter may resemble subsurface conditions in summer, and may be similar in relation to maximum light intensity and day length (i.e. amount of time PE populations are exposed to light) (Chapter 3, Figure 3.4C, p.79).

Ecological populations occupied intervals closer to the mat surface in EZ2 compared to EZ1 at MS (Chapter 3, Figure 3.3, p.78), possibly because they require higher light intensities to reach the compensation point where net photosynthesis can

occur at higher temperatures (Meeks and Castenholz, 1971; Allewalt et al, 2006). At lower light intensities in December, PEs A1 and A5 increased in relative abundance in EZ3, possibly indicating that they are surface-oriented populations in months with reduced light intensity in order to be able to reach the compensation point where net photosynthesis can occur at higher temperatures. However, B′-like PEs do not increase in relative abundance at warmer temperatures in months with reduced light intensity, possibly because they are incapable of crossing the 63/64°C ecotone boundary. In contrast, populations at different temperatures and depths could be young ecological species that could not be separately demarcated using the *psaA* locus. Further experiments are needed to confirm the seasonal differences observed and to assess vertical distributions in different seasons.

psaA Sequence and Putative Ecotype
Variation in MS and OS between 1989 and 2009

Comparisons among samples collected over a period of 20 years was complicated by the fact that collections were made in different springs, in different months and at different temperatures. Thus, as outlined in Figure 5.4, it was necessary to separate these samples by spring, month of collection and EZ, as inferred from comparison of PE distributions to temperature-defined samples collected in 2006 and 2008 (Chapters 2 and 3). MS samples were collected from EZ1 (between ≤60 and 63°C) and EZ2 (between 64 and 65°C) in 1996, 1999, 2006, 2007, 2008 and 2009, and OS samples were collected from the same EZs in 1989, 1991, 1996, 1999 and 2005 (Table 5.2). A total of 30,595 sequences (average of 2,353; standard deviation of 492) were analyzed yielding 118

HFSs. In Figures 5.5 and 5.6, PEs demarcated by ES are indicated with vertical bars; the percent of each HFS (and surrounding variation) in the total sample (numbers left of vertical bars), and the PE percent population ((PE sequences in sample / all sequences in sample) x100)) of the total sample (numbers to the right of vertical bars) are also reported. Separate sampling by spring and EZ allowed me to better observe within-population changes that had occurred within clusters of sequences predicted to be ecologically distinct over the period of time studied, as well as total PE population fluctuations.

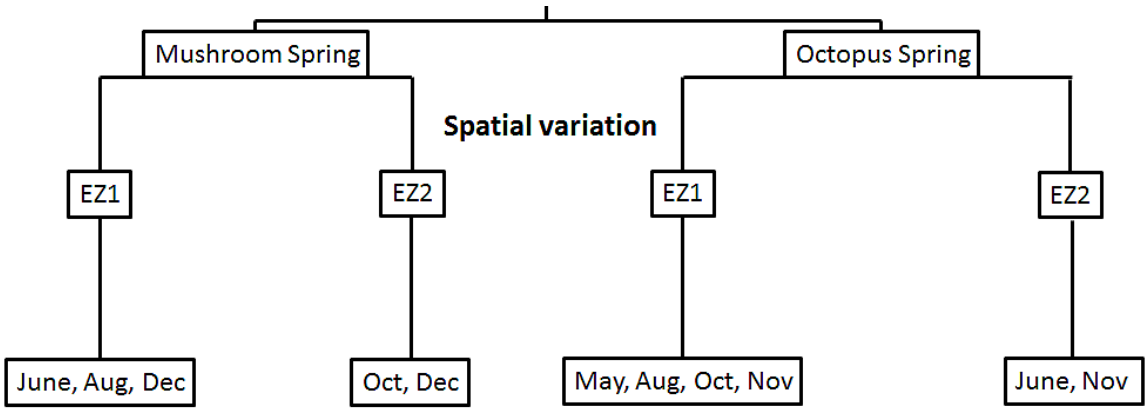


Figure 5.4. Pattern of analyses in Octopus Spring and Mushroom Spring (MS) samples based on collection in different ecological zones (EZs) (EZ1, EZ2 and EZ3) and months that correspond to variation presented in Figures 5.5 and 5.6.

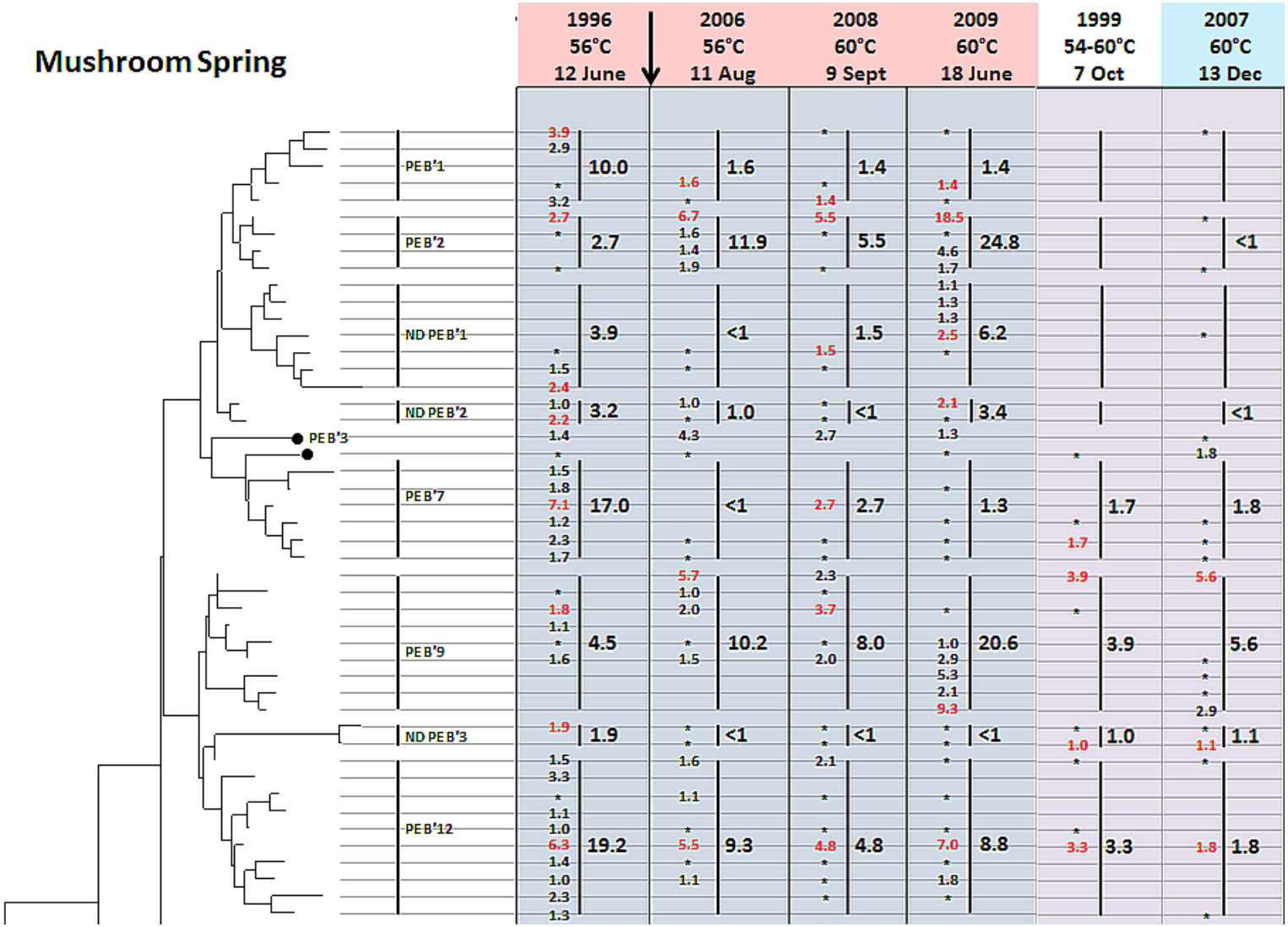
PE variation within ecotone boundaries generally matched the diversity observed in 2006 cloning and sequencing studies (Chapter 2; Becraft et al., 2011) and 2008/2009 Ti454-barcoding studies (Chapter 3; Becraft et al., submitted). Reduced sequence length (from 324 to 302 base pairs) caused ES to lump together some A-like PE populations previously shown to be ecologically distinct (e.g. PE A14 is lumped with PE A1) (see

Chapter 3, Tables 1 and 2, pp.87 and 88). Additional PEs were demarcated in this analysis that were not identified in 2006 and 2008 studies when samples from previous years were included, and are labeled as 'newly demarcated' (ND) PE #; where # is numbered sequentially from 1 to n, and correspond to PE designations in Chapter 4.

Mushroom Spring Phylogenetic Analysis and Population Dynamics.

ES predicted a total of 13 B′-like, 2 A′-like and 14 A-like PEs at Mushroom Spring (Figure 5.5). ES calculated periodic section and ecotype formation rates of 11.0 and 0.61 events/nucleotide substitution, respectively, indicating a much higher rate of periodic selection than ecotype formation. Samples from 1999 and 2007, which were likely taken at higher average temperatures, are representative of diversity in EZ2. Samples from 1996, 2006, 2008 and 2009 were representative of EZ1; however, the 2006 and 2009 samples appear to be from a cooler average temperature where PE B′2 is more relatively abundant compared to samples taken closer to the EZ2 ecotone boundary, and there are lower numbers of A-like HFSs (see Chapter 3, Appendix B Figure B3). There were five newly demarcated B′-like PEs, as well as one newly demarcated A-like PE, that were not based on single HFSs. New PEs not sampled in more recent years (Chapters 2 and 3) could be populations that have since declined in relative abundance and/or that were not detected in previous samples (e.g. 2006 cloning and sequencing and 2008 Ti454-barcoding studies (Chapters 2 and 3)).

MS provided the best opportunity to analyze the population dynamics of PEs over a period of years, as samples were all from summer months when light was relatively


Mushroom Spring
1996
56°C
12 June
2006
56°C
11 Aug
2008
60°C
9 Sept
2009
60°C
18 June
1999
54-60°C
7 Oct
2007
60°C
13 Dec
PEB'1
PEB'2
ND PEB'1
ND PEB'2
PEB'3
PEB'7
PEB'9
ND PEB'3
PEB'12
10.0
1.6
1.4
1.4
2.7
11.9
5.5
24.8
<1
3.9
<1
1.5
6.2
3.2
1.0
<1
3.4
<1
17.0
<1
2.7
1.3
1.7
1.8
4.5
10.2
8.0
20.6
3.9
5.6
1.9
<1
<1
<1
1.0
1.1
19.2
9.3
4.8
8.8
3.3
1.8

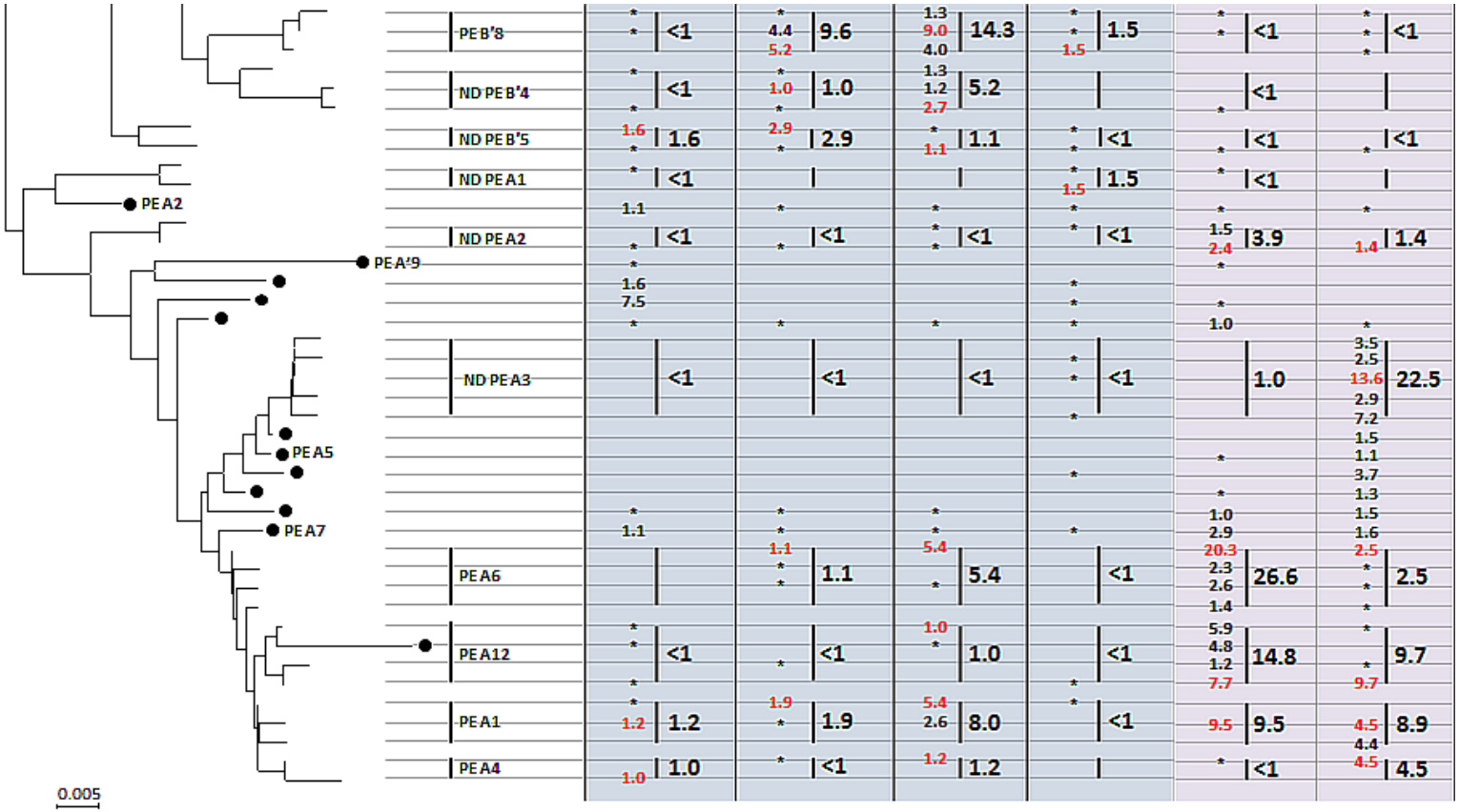


Figure 5.5. High-frequency sequence (HFS) phylogeny of A′-like, A-like and B′-like *psaA* sequences that were ≥1% in a given sample at Mushroom Spring from 1996 to 2009. Shading represents different ecological zones (EZs) (blue = EZ1; purple = EZ2); color shading on top represents season (red = summer; blue = winter; white = fall/spring). Vertical bars indicate putative ecotypes (PEs) demarcated by Ecotype Simulation; percent of each HFS in the sample is indicated to the left of the vertical bars, and PE percent population is shown to the right of the vertical bars (summed HFSs). Newly demarcated PEs are labeled ND PE. PE dominant variants are colored red. Dots represent singleton HFS PEs. Arrow indicates a shift in the main flow path from the source. Scale bar represents 0.005 substitutions per site.

similar so that seasonal variation should have contributed little to the sample diversity (Figure 5.5; EZ1, 1996-2009) (also see Figure 5.2, inset box). In EZ1, PEs B′1, B′7 and B′12 were higher in relative abundance in 1996, and lower in relative abundance in all other years analyzed (ranging from 10-1.4%, 17-<1% and 19.2-4.8%, respectively). In contrast, PEs B′2, B′8 and B′9 were higher in relative abundance in 2006 and afterward, and were lower in relative abundance in 1996 (ranging from 2.7-24.8%, <1-14.3% and 4.5-20.6%, respectively). ND PE B′1 was lower in relative abundance in 2006 and 2008, though was higher in relative abundance in 1996 and 2009 (<1-6.2%).

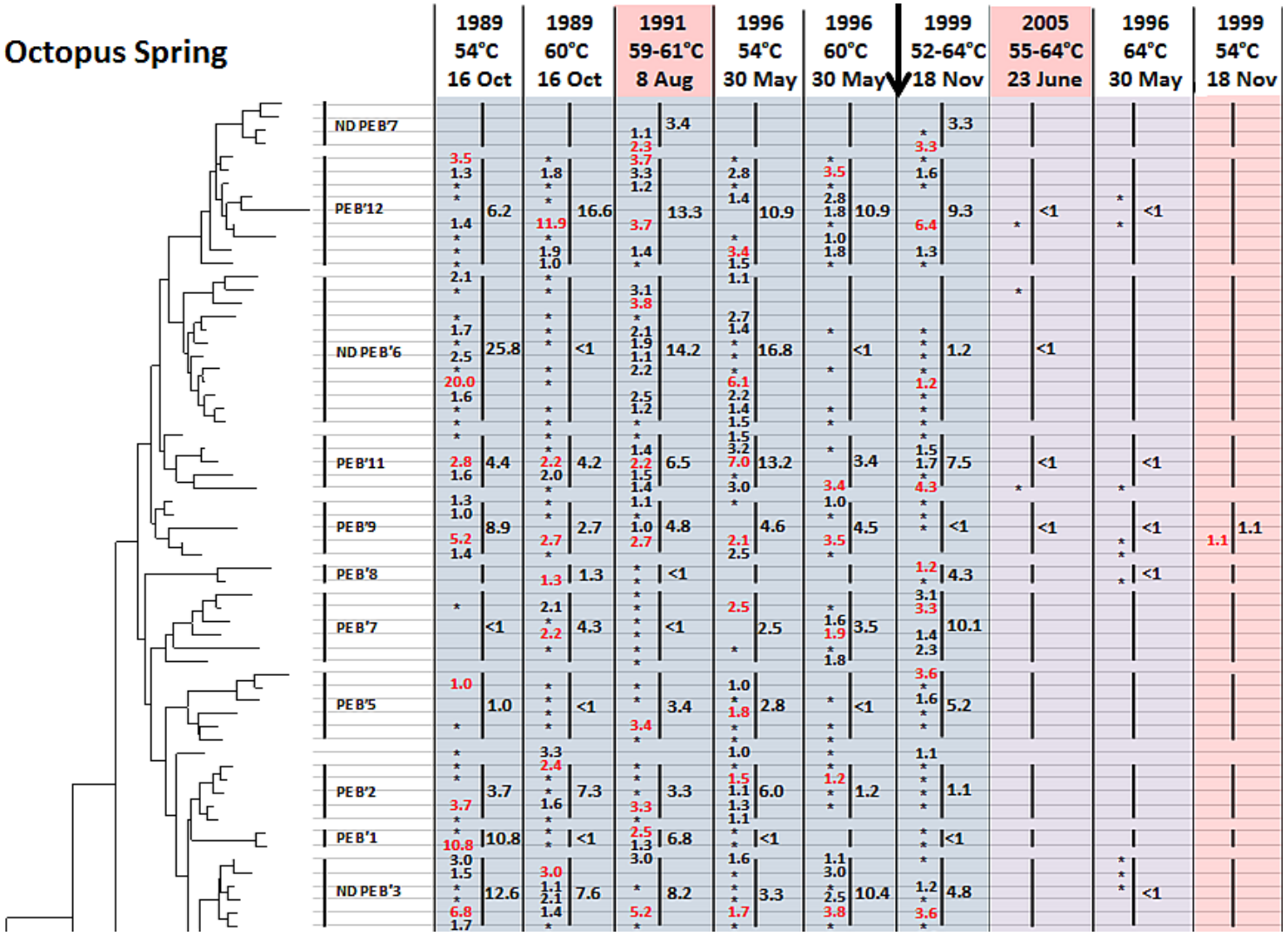
In 2005, the main flow path shifted directions ~180° at a point ~5 m downstream from the source pool at MS (Figure 5.5, arrow). This change of direction in the main flow path could have occurred as a result of the damming of lower temperature areas or upstream depressions creating new outflows (Ward, 2012), which would have the greatest impact on the downstream populations (e.g. B′-like ecotypes). This environmental disturbance would have provided the opportunity for growth on the uncolonized surface. PE B′7 might be an example of a population that was displaced by other populations after the disturbance event (Figure 5.5). In contrast, PE B′2 might an example of a population that was able to take advantage of the newly forming mat environment after disturbance. However, this argument is weakened by the absence of data in the ten years between sample collections (1996 to 2006).

In EZ2, A-like PE populations appeared to shift significantly between 1999 and 2007; PEs A4 and ND PE A3 were higher in relative abundance (ranging from <1-4.5% and 1-22.5%, respectively), while PEs A6 and A12 were lower in relative abundance

(ranging from 26.6-2.5% and 14.8-9.7%, respectively), in 2007 compared to 1999. The 1999 and 2007 samples were collected in different seasons, which may have contributed to the observed differences in A-like PE relative abundances.

Octopus Spring Phylogenetic Analysis and Population Dynamics. ES predicted 16 B′-like PEs, 6 A′-like PEs, and 8 A-like PEs at OS (Figure 5.6). ES calculated a periodic section rate of 75.0, and an ecotype formation rate of 0.85 events/nucleotide substitution, resulting in an even higher ratio of periodic selection/ecotype formation than was observed in the MS year-to-year samples. Samples from 1999 and 2005 correspond to EZ2 and all other OS samples contained PE diversity representative of EZ1, though samples were collected in various seasons (spring, summer and fall) (Figure 5.2). In OS there were 5 newly demarcated B′-like PEs (including 3 new PEs not identified in MS), 1 newly demarcated A-like PE and 4 newly demarcated A′-like PEs that were not singleton HFSs.

PEs were observed to change in relative abundance in EZ1 at OS from 1989 to 1999 in multiple seasons, though the samples from October 1989 and November 1999 were both collected in fall at similar temperatures. Additionally, the 60°C August 1991 sample and the May 1996 sample had similar average light intensities during the time of collection (Figure 5.2, arrows). In EZ1, ND PE B′3, ND PE B′6 and PEs B′1 and B′9 were higher in relative abundance in 1989 (54°C), and lower in relative abundance in November 1999 (54°C) (ranging from 12.6-4.8%, 25.8-1.2%, 10.8-<1% and 8.9-<1%, respectively). PEs ND B′1, B′5, B′7 and B′8 were lower in relative abundance in 1989


Octopus Spring
1989 54°C 16 Oct
1989 60°C 16 Oct
1991 59-61°C 8 Aug
1996 54°C 30 May
1996 60°C 30 May
1999 52-64°C 18 Nov
2005 55-64°C 23 June
1996 64°C 30 May
1999 54°C 18 Nov
ND PEB'7
PEB'12
ND PEB'6
PEB'11
PEB'9
PEB'8
PEB'7
PEB'5
PEB'2
PEB'1
ND PEB'3

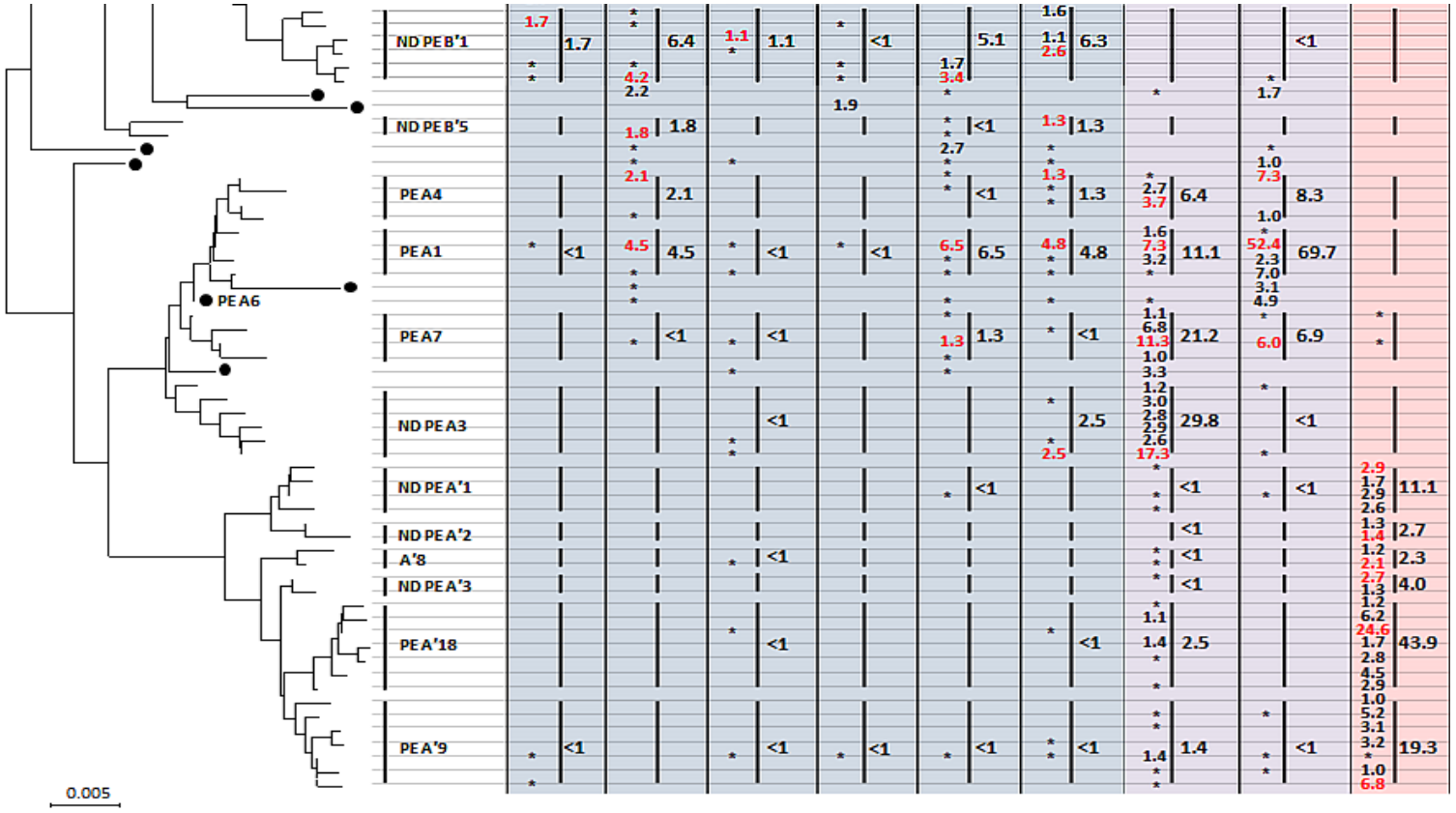


Figure 5.6. High-frequency sequence (HFS) phylogeny of A′-like, A-like and B′-like *psaA* sequences that were ≥1% in a given sample at Octopus Spring from 1996 to 2009. Shading represents different ecological zones (EZs) (blue = EZ1; purple = EZ2; red = EZ3); color shading on top represents season (red = summer; blue = winter; white = fall/spring). Vertical bars indicate putative ecotypes (PEs) demarcated by Ecotype Simulation; percent of each HFS (and surrounding variation) in the sample is indicated to the left of the vertical bars, and PE percent population is shown to the right of the vertical bars (summed HFSs). Newly demarcated PEs are labeled ND PE. PE dominant variants are colored red. Dots represent singleton HFS PEs. Arrow indicates a major hail storm that destroyed the mat. Scale bar represents 0.005 substitutions per site.

(54°C), and higher in relative abundance in subsequent years including 1999 (ranging from 1.7-6.3%, 1-5.2%, <1-10.1% and 0-4.3%, respectively). Additionally, PEs ND B′1 was higher in relative abundance in May (60°C) compared to August 1991 (ranging from 5.1-1.1%), and NDB′6 was lower in relative abundance (ranging from <1-14.2%), providing further evidence that these PEs were different in relative abundance as a result of yearly population dynamics.

In 1996, a hail storm heavily damaged the mat at OS (Figure 5.6, arrow). This environmental disturbance may have provided the opportunity for the colonization of newly forming mat at damaged sites between 1996 and 1999. PE B′9 might be an example of a population that was displaced by other populations after the disturbance event. In contrast, PE B′7 might be an example of a population that was able to take advantage of the newly forming mat environment after disturbance.

In EZ2, PE A1 was lower in relative abundance in 1999 compared to 2005 (ranging from 11.1-69.7%), and ND PE A3 and PE A7 were higher in relative abundance (ranging from 29.8-<1% and 22.1-6.9%, respectively). In 2005, ND PE A3 was <1% of the sample and PE A7 was also in lower relative abundance, while PEs A1 and A12 were higher in relative abundance, with PE A1 equaling ~70% of the total sample diversity. Similar to MS, total A-like diversity changed between 1999 to 2005, but these samples were also collected in different seasons (late fall and summer), and the observed variation in PE relative abundance is possibly due to seasonal population dynamics.

Population Genetics of Putative Ecotypes at Mushroom Spring and Octopus Spring from 1989 to 2009. Figures 5.5 and 5.6 show the DV (red numbers) and other HFSs contained within PEs changing yearly. The observed changes in the DV and surrounding HFSs are what would be expected if periodic selection events and/or neutral genetic drift were acting upon the variation predicted to be contained within PE populations. Additionally, this could indicate that there are multiple ecotypes cryptically contained within a PE, and their DVs are changing independently of one another. PEs B′2 and B′12 in MS, and PEs A1, A7 and B′9 in OS, maintained the same dominant variant across all years studied where the percent population was >1%, and there were only minor fluctuations in the less abundant HFSs. This could indicate that these populations have not experienced a periodic selection event over the time period analyzed, even in response to recolonization after environmental disturbance. In other cases, PE DVs changed between 1996 and 2009; PE DVs sometimes shifted only once over the period of time analyzed (e.g. B′11 in OS after the 1996 hail storm), and some PE DVs also changed nearly every year sampled (e.g. PE B′12 at OS). These results are consistent with the high rate of periodic selection calculated by ES. While difficult to conclude that changes in the relative abundance of DVs and HFSs within a PE are due to periodic selection events and/or genetic drift, the patterns are what would be expected of ecological populations responding to these evolutionary forces over time. It is interesting to note that the DVs of predominant PEs did not change seasonally in 2011.

Table 5.3 summarizes the number of times a DV changed within a PE clade in MS and OS. MS PEs shifted DVs 17 times over 13 years, and OS PEs shifted DVs 35 times over

16 years. The basis for this difference is unclear, though there is more environmental variation in Octopus Spring due to flow surges, and the types of disturbances in MS and OS were different, and these factors might have been involved. For instance, in MS, the flow-path redirection of 180 degrees would have necessitated colonization of uncolonized surfaces, whereas in OS, recolonization after a hail storm might have involved reestablishment atop damaged mat where PEs that were present before the disturbance may have also been present after the disturbance.

Table 5.3. Number of dominant variant (DV) changes for abundant putative ecotypes (PEs) with >1 high-frequency sequence (HFS), and that totals ≥1% of any sample.

	Number of DV shifts	
PEs	Mushroom Spring [a]	Octopus Spring [b]
B′1	3	2
B′2	0	4
B′5	1	4
B′7	1	4
B′8	2	1
B′9	4	0
B′11		0
B′12	0	4
NDB′1	3	4
NDB′2	1	
NDB′3	1	3
NDB′4	1	
NDB′5		1
NDB′6		3
NDB′7		3
A1	0	0
A4	0	1
A6	0	
A7		0
A12	0	
NDA1	0	
NDA2	0	
NDA3(A5)	0	0
TOTAL	17	35

[a] Mushroom Spring samples were collected over a 13 year period in ecological zone (EZ) 1, and over a 8 year period in EZ2.

[b] Octopus Spring samples were collected over a 10 year period in EZ1, and over a 6 year period in EZ2.

Additionally, while there were fewer EZ2 samples, and the periods of time between sampling were shorter, A-like populations very rarely shifted DVs within a PE despite large fluctuations of PE relative abundances between seasons. This could indicate that A-like populations are more stable than B′-like PEs, or alternatively, this could be an effect of less HFS diversity contained within A-like populations at the *psaA* region analyzed (i.e. A-like PEs are younger ecotypes), and reduced within-PE diversity is essentially hiding the DVs changing within A-like PEs over time.

The percent of DVs and HFSs within-PEs was calculated to observe within-ecotype population structures over the period of time studied (see Chapter 8, Table 8.1, p.216). While there was year-to-year variation in the number of HFSs contained within PEs, it usually correlated with the total relative abundance of the PE in the mat. Within-ecotype genetic structure is examined in more detail in Chapter 8.

Conclusions

Ferris and Ward (1997) suggested that 16S rRNA genotypes remained constant seasonally in OS, though 16S rRNA sequences lacked the molecular resolution to track ecological populations. When a more highly-resolving protein-encoding locus was analyzed, there was a higher relative abundance of A-like PEs in EZ1 and EZ3 in December. PEs typically found in EZ3 (A′-types) were lower in relative abundance, and PEs typically only relatively abundant in EZ2 (PEs A1 and A5), were higher in relative abundance. Also, A-like PEs A4 and A5 were higher in relative abundance in EZ1, where B′-like PEs are typically more predominant in the summer. This could indicate that PEs

A4 and A5 are better adapted to winter light conditions, and provides evidence that *psaA*-defined PEs shift in their relative abundances seasonally. Similar results were also observed for *Synechococcus* PEs in the water flowing over the mat (see Chapter 6, Figure 6.6, p.171). PE seasonal population dynamics provided evidence that ecotone boundaries are not necessarily determined by temperature alone, but by a combination of environmental parameters (e.g. light and temperature), though more studies are needed to analyze the seasonal distributions and population dynamics of *Synechococcus* ecotypes.

While difficult to compare MS and OS directly, PE relative abundances in the OS 1999 (November; late fall) and MS 2007 (December; winter) samples shown in Figures 5.5 and 5.6 corresponded to differences in PE relative abundance observed in the December 2011 samples (Figure 5.3). PE A5 was observed to be higher in relative abundance in winter months in 2011 (Figure 5.3 and Chapter 6, Figure 6.6, p.171). While demarcated separately in this analysis, ND PE A3 branches close to the DV of PE A5 (Figure 5.5). The resulting change of DV could be the result of periodic selection and/or genetic drift acting upon PE A5 between 2007 and 2011. In MS, PE A4 and ND PE A3 were higher in relative abundance in the December 2007 sample; in OS, PE A7 and ND PE A3 were higher in relative abundance in November 1999.

PE relative abundances fluctuated between samples over the period of time analyzed in both MS and OS, possibly because of seasonal differences. However, subsets of samples that were taken in the same EZ and season showed that the relative abundances of PE populations in both springs changed yearly. In MS, the 1996-2009 EZ1 samples were all collected in summer, and two subsets of samples at OS were collected

during times of the year with similar average light intensities (October/November and August/May), which allowed for the differentiation between seasonal changes and long-term changes. Some populations decreased in relative abundance to ≤1% of the sample in more recent years (e.g. ND PE B′6 in Figure 5.5), though no PEs went completely extinct within an EZ. PEs also increased in relative abundance from <1% to predominance (e.g. PE B′2 in MS).

Changes in B′-like PE relative abundances corresponded to historical environmental disturbances. In 2005, the main flow path shifted directions from the source pool at MS (Figure 5.5, arrow), and in 1996, a hail storm heavily damaged the mat at OS (Figure 5.6, arrow). Additionally, the relative abundances of PE populations in the OS 1989 sample could have been influenced by the Yellowstone fires of 1988, which possibly influenced carbon and nitrogen levels in water sources (Romme et al., 2011). The minimal seasonal variation of samples collected from the same EZ and time of year supports the inference that the decreased relative abundance of prominent PEs, and the increased relative abundance of low-abundance PEs were the result of year-to-year population dynamics in these local communities, possibly in response to environmental disturbances.

Periodic selection rates were extremely high compared to all other analyses (see Chapters 3 and 4), which could have been the result of including within-PE variation from multiple years. DVs and HFSs within PEs did change, possibly as a result of the forces of periodic selection and/or genetic drift acting upon these populations. The high rate of periodic selection calculated by ES in MS and OS over the period of time

analyzed would predictably result in the more frequent changing of PE DVs and HFSs temporally, as was observed for many PEs in Figures 5.5 and 5.6 (Table 5.3).

Differences in the stability of DVs between springs could have been the result of PEs initially colonizing newly forming mat after environmental disturbance under different circumstances (colonizing a newly forming effluent channel in MS vs. sporadic hail storm damage in OS that may have left some populations undisturbed along the flow path). The genetic structure of PEs differed, indicating that PEs changed differently from one another seasonally and yearly. In some cases PE DVs remained constant, and other PE DVs changed nearly every year sampled over the period of time analyzed. These results indicate that *Synechococcus* PEs are evolving separately from one another, which would be expected of ecological populations occupying unique environmental niches or separated by geographic distance (e.g. MS vs. OS, see above).

CHAPTER 6

COLONIZATION OF DISTURBED MUSHROOM SPRING MAT BY *SYNECHOCOCCUS* ECOLOGICAL POPULATIONS UNDER DIFFERENT LIGHT CONDITIONS

Introduction

Hot spring microbial mats in Yellowstone National Park are at the mercy of nature, often experiencing major natural disturbances such as hail storms, flow path changes, source pool temperature fluctuations, forest fires and damage from wildlife. To better understand and preserve these natural resources, it is important to understand the impact these disturbances can have on microbial communities, the ecology of recovering mats, and the role of cell dispersal and adaptation on initial colonization (Morin, 2011; Chapter 4, Figure 4.1, p.94). The following chapter addresses the effects of mat disturbance on the *Synechococcus* cyanobacterial populations in recovering mats, the possible role of adaptation to different light environments during recovery, and the cellular content of water that flows above the mat surface as a possible factor contributing to dispersal and mat recolonization.

Microbial succession is the sequential development of microbial populations on uncolonized surfaces or after an environmental perturbation. The process of succession has been studied in many diverse microbial ecosystems to, for instance, (i) understand the effects of climate change (Kaštovská et al., 2005), (ii) analyze fermentation processes in human diets (Coppola et al., 2000), (iii) examine colonization of the human mouth after

extensive dental work (van Winkelhof et al., 1988), (iv) discern the patterns of succession during litter decomposition in forests (Torres et al., 2005), (v) better understand large-scale composting processes (Ishii et al., 2000), and (vi) better direct reclamation efforts of human-contaminated sites (Fries et al., 1997; Mueller et al., 2011).

It is important to understand succession in Yellowstone hot spring microbial communities to assess whether anthropogenic or natural disturbances affect colonization of newly forming mats. Microbial mats in Yellowstone alkaline siliceous hot springs are well-studied systems that allow for the opportunity to conduct fundamental studies of disturbance ecology and mechanisms of cell dispersal (Ward et al., 1998; Ward and Castenholz, 2000). As described in the previous chapters, we have a good understanding of *Synechococcus* ecological species inhabiting Mushroom Spring (MS), their distributions and their responses to environmental perturbations. As a consequence, it is a system in which to better understand the involvement of processes related to dispersal and initial colonization, and to the role of adaption in community reassembly.

The ~1 mm-thick top green layer of the microbial mats at in Yellowstone alkaline siliceous hot springs have *Synechococcus* population densities as high as 10^{10} cells per cm^3 (Brock, 1978). Brock and Brock (1969) studied the recovery of a hot spring microbial mat following a hail storm by analyzing *Synechococcus* cell density with time after disturbance. These investigators found that there was partial mat recovery after only 10-17 days (from ~10^4 to 10^6 cells/ml), and further recovery after 152 days (~10^8 cells/ml). With the advent of molecular methods, Ferris et al. (1997) used denaturing

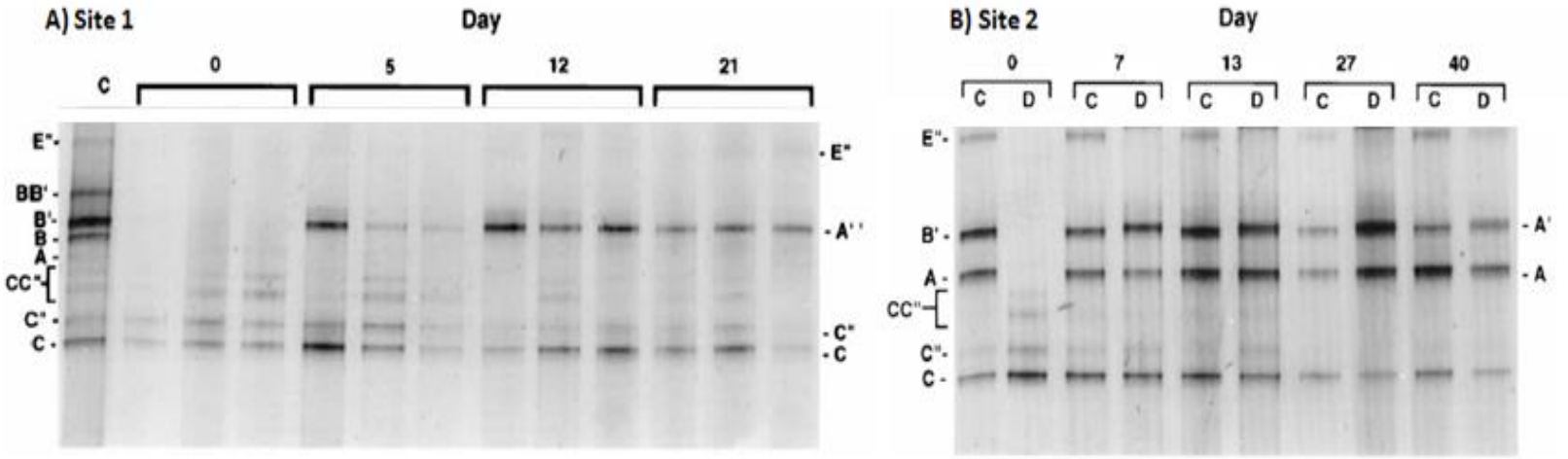


Figure 6.1. Denaturing gradient gel electrophoresis (DGGE) profiles of 16S rRNA variants found with time after physical disturbance by removal of the top green layer of the Octopus Spring microbial mat. (A) Site 1 (55 to 62°C) mat sample control (lane C), and triplicate profiles immediately (day 0) and 5, 12, and 21 days after disturbance. (B) Site 2 (58 to 62°C) mat sample control (lanes C) and disturbed (lanes D) samples immediately (day 0) and 7, 13, 27, and 40 days after disturbance. Single letters indicate 16S rRNA sequence types; double letters indicate heteroduplex bands (modified from Ferris et al. 1997).

gradient gel electrophoresis (DGGE) analysis of 16S rRNA sequences to investigate the recovery of microbial community structure after removal of the top green mat layer containing *Synechococcus* cells at two different temperature-defined sites (Site 1 (55 to 62°C) and Site 2 (58 to 62°C)) in Octopus Spring (OS) (Figure 6.1). At these sites, which are in ecological zone (EZ) 1 (a region of the mat with unique genetic diversity located between ≤58 to 63-64°C; also see Chapter 3, Figure 3.2, p.77), the *Synechococcus* A, B′, and B 16S rRNA genotypes normally present in the upper green mat layer (Ferris et al., 1996) were removed by scraping (compare lanes C and 0 in Figure 6.1A, and day 0 C and D in Figure 6.1B). During mat recovery, genotypes previously identified at that location mostly remained absent for the duration of the study, although A and B′ 16S rRNA genotypes did begin to return in later stages at site 2 (Figure 6.1B). The A′ 16S rRNA genotype, which is normally located at higher temperatures (EZ3; region of the mat with distinct genetic diversity located at ≥66°C; Chapter 3, Figure 3.2, p.77), was detected after disturbance at Site 2, and a previously unidentified 16S rRNA genotype related to the A′ genotype (A′′′) was observed after disturbance at Site 1. Thus, it appeared that dispersal of cells from upstream populations might be important in the early successional phase. There was also concern that the 16S rRNA locus may have been too conserved to discern *Synechococcus* species (Melendrez et al., 2011), as it has been shown to lump many different ecological species that were subsequently shown to have different ecologies (Becraft et al., 2011; Becraft et al., submitted). Furthermore, Ferris et al., (1997) did not examine colonization until five to seven days following disturbance (Figure 6.1). To determine whether the newly colonizing cells were different ecological

species, the more highly-resolving genetic marker, *psaA*, used in the previous chapters of this thesis, was analyzed in samples collected at daily intervals after disturbance. In some cases, disturbed environments were altered to assess whether ecological factors, specifically removal of UV light or reduction of light intensity, played a role in recolonization during this period.

Because A′-like cells had been observed to recolonize disturbed sites (Ferris et al., 1997), I also investigated *Synechococcus* genotypes in the flow. The microbial mats of MS and OS have different hydrodynamic properties and flow rates, which in turn can influence sloughing of cells from the mat surface, cell dispersal and colonization after mat disturbance. PE populations identified in MS were shown to have unique distributions along vertical gradients where light and oxygen are known to vary (Chapter 2, Figure 2.4 and Chapter 3, Figure 3.3, pp. 49 and 78). Populations located upstream and surface-oriented populations should theoretically make greater contributions to the cell content in the water overflowing a downstream site than the populations at the downstream sites due to cells sloughing from the mat surface and accumulated dispersal along the flow path. Hence, they should have more influence on initial colonization at downstream sites. Additionally, populations better adapted to different light environments should be able to colonize newly forming mat under different light conditions. To test these hypotheses about the roles of cell dispersal and population-specific adaptation in mat recolonization in these springs, the PE diversity in cells filtered from water samples collected at different temperature-defined sites was analyzed along the effluent flow path, and during different seasons.

Materials and Methods

Fieldwork

Disturbance and Light Alteration. A 58.6 to 62.2°C (EZ1) site in the microbial mat of MS was covered with a wire mesh screen supported by a wooden platform placed ~1 to 2 cm above the water surface to protect the experimental area from possible hail damage (Figure 6.2). The area underneath the screen was divided into 6 separate sections, each of which corresponded to a different environmental condition: (i) undisturbed mat (control), (ii) scraped (removal of top green layer), (iii) UV-blocked, (iv) ~92% light reduction, (v) scraped and UV-blocked and (vi) scraped and ~92% light reduction. To examine the effects of light alteration, one section was covered with 4 layers of stretched muslin to reduce downwelling photon irradiance by ~92%, and another section was covered with UV-blocking Plexiglas (UF-5; 3.2-mm thick; Plexiglass: Autohaas N. Am., Philadelphia, PA, USA) as described in Chapter 3 Material and Methods. To analyze the recolonization of newly forming mat after disturbance, the upper green mat layer was completely removed in 3 separate 7 x 10 cm areas using a bent spatula, with minimal removal of the red undermat. For the light alteration experiments, the scraped areas were located beneath the center of each light cover. Samples were collected in duplicate from each of the six sections by David Ward from 21 to 26 July 1996 using a #4 cork borer (38.5 mm^2) every day between 1200 and 1800 h. Samples were immediately frozen on dry ice (-20°C) in the field and kept frozen at -80°C until analysis in 2011. The 1996

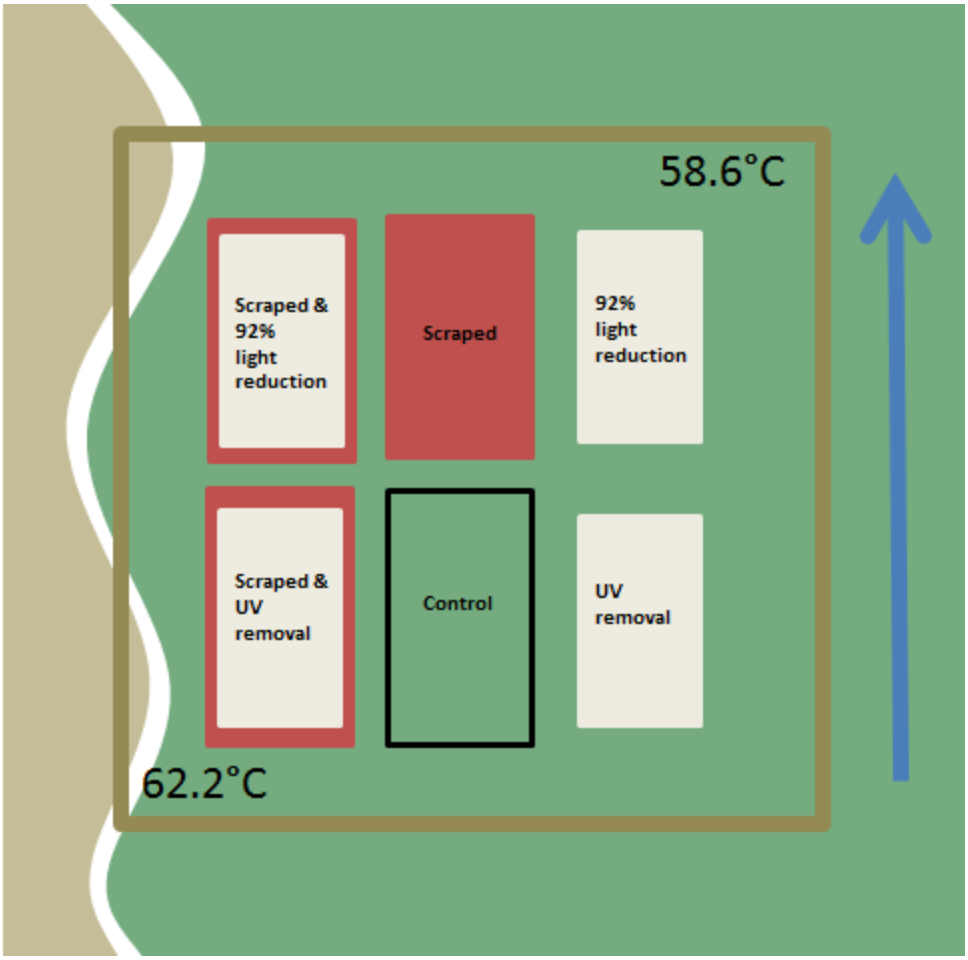


Figure 6.2. Experimental design for 1996 disturbance and perturbation studies; schematic diagram showing the orientation of the six experimental sections under the protective screen. Blue arrow indicates direction of water flow.

experimental design included studies of the effect of light and temperature alteration similar to those reported in Chapter 3. In order to focus on colonization after disturbance, results of these perturbation analyses without disturbance (i.e. mat scraping) are reported in Appendix D, Sections I and II; Appendix Tables D2-D4).

Cells in Flow. Water samples (which contained 10^3 to 10^4 *Synechococcus* cells/ml) were collected by collaborating doctoral student Shane Nowack (Isaac Klapper

laboratory, MSU Dept. of Mathematics and Temple Univ. Dept. of Mathematics) on 2 June 2011, and 15 December 2011, at 5 sites along the effluent flow path at MS and OS mat. Taking care to avoid mat material, 100 ml of water was collected ~1 cm above the mat surface from each site and was passed through a 0.2 µm Whatman Millipore filter, which was immediately frozen on dry ice (-20°C). Temperature data were recorded using the DS1922T Temperature Logger iButton with 8KB memory (Maxim Integrated, San Jose, CA) as described in Chapter 5 Materials and Methods (Table 6.1).

Table 6.1. Average, minimum and maximum temperatures taken over a 24 hour period at Octopus Spring in June 2011. Sample sites progress downstream along the effluent flow path from site 0 (just downstream from the source) to site 4. Ambient temperature is also reported. Temperature data for Mushroom Spring are presented in Chapter 5, Table 5.1.

	Site 0	Site 1	Site 2	Site 3	Site 4	Ambient
June 6 OS[a]						
24 hr-avg	¥	66.2°C	57.3°C	54.7°C	45°C	10.26°C
Max		74.3°C	67.1°C	59.4°C	48.2°C	25.66°C
Min		60.3°C	52°C	49.5°C	41.1°C	2.65°C

[a] OS, Octopus Spring

¥ Missing data are due to failed temperature data-loggers, usually because of prolonged exposure to high temperatures, and point measurements taken at the time of collection are reported in these cases.

Molecular Methods

Cells filtered from water samples were lysed directly from the filters by bead-beating as described in Chapter 2 Materials and Methods. DNA was extracted from water-derived and mat samples, PCR amplified at the *psaA* locus, and sequenced with the 'centerfowardprimer' using Ti454-barcoding technology at either the J. Craig Venter Institute (mat samples) as described in Chapter 3, or at Research and Testing Laboratory, Lubbock, TX (filtered water samples), as described in Chapter 5 Materials and Methods.

Sequences were trimmed to 302 base pairs to obtain the maximum number of sequences, cleaned and analyzed to identify high-frequency sequences (HFSs; sequences equaling ≥50 identical copies in all samples combined) and associated low-frequency sequences (LFSs; sequences equaling <50 copies in all samples combined) as described in Chapter 3 Materials and Methods (Becraft et. al., submitted). Replicates at some time points were lost due to failed PCR or sequencing reactions.

A-like (fine-scale demarcation) and B′-like (conservative demarcation) *Synechococcus* PEs were demarcated separately as described in Chapter 3 Materials and Methods (also see Appendix B, Section I). A′-like diversity was greatly increased compared to previous studies (Chapter 3), possibly because of the inclusion of populations that inhabited the source pool and water flowing above higher-temperature mat regions. Since the A′-like lineage was more divergent than the A-like lineage, A′ PEs were demarcated using the conservative approach. Newly identified PEs were labeled 'newly demarcated' (ND) PE #, where # is labeled sequentially from 1 to n. ND PEs corresponded to those identified in Chapters 4 and 5 unless discovered in this disturbance experiment. Reduced sequence length caused low-abundance PEs B′1, B′3 and PE A′18 to be demarcated within PEs B′4, B′6 and PE A′9, respectively. Percent population of each PE in a sample was calculated by dividing the total number of sequences within a PE detected in a sample (PE sequences = DV sequence + all other HFSs + all LFSs) by the total number of sequences in that sample ((PE sequences in sample / all sequences in sample) x100). Only PEs that were ≥5% in a sample are shown in the main text, but results for all PEs are presented in Appendix D Tables D2 and D5-D7.

Results and Discussion

The 1996 disturbance experiment yielded 178,709 total sequences (average of 1,477 per sample; standard deviation of 537) from the analysis of 121 samples, and 340 HFSs were identified from these sequences. The high number of HFSs was due to previously unidentified A′-like PEs demarcated during mat recovery studies, and the increased total number of samples and sequences analyzed. Most newly identified A and B′-like HFSs were contained within PE clades that had been observed in previous studies (Chapter 3, Figure 3.1, p.75), as opposed to defining new PEs. ES predicted a total of 40 PEs (18 B′-like, 12 A-like and 10 A′-like), though A-like PEs were low in relative abundance in nearly all samples analyzed. ES calculated a periodic section rate of 0.8 events/nucleotide substitution and an ecotype formation rate of 0.4 events/nucleotide substitution, indicating a higher rate of periodic selection than ecotype formation (2:1 ratio of periodic selection/ecotype formation events).

Site Characterization and Immediate Effect of Disturbance

As shown in Figure 6.3A, PE diversity of the undisturbed experimental site mostly matched the diversity observed in EZ1, where B′-like PEs are the predominant populations (Chapter 3; Figure 3.2, p.77). However, the 1996 disturbance site appeared to be at a slightly cooler average temperature than EZ1 samples collected at sites closer to the 63-64°C ecotone boundary in 2006/2008 (see Chapter 2, Figure 2.3 and Chapter 3, Figure 3.2, pp. 47 and 77). This inference was based on the greater relative abundance of

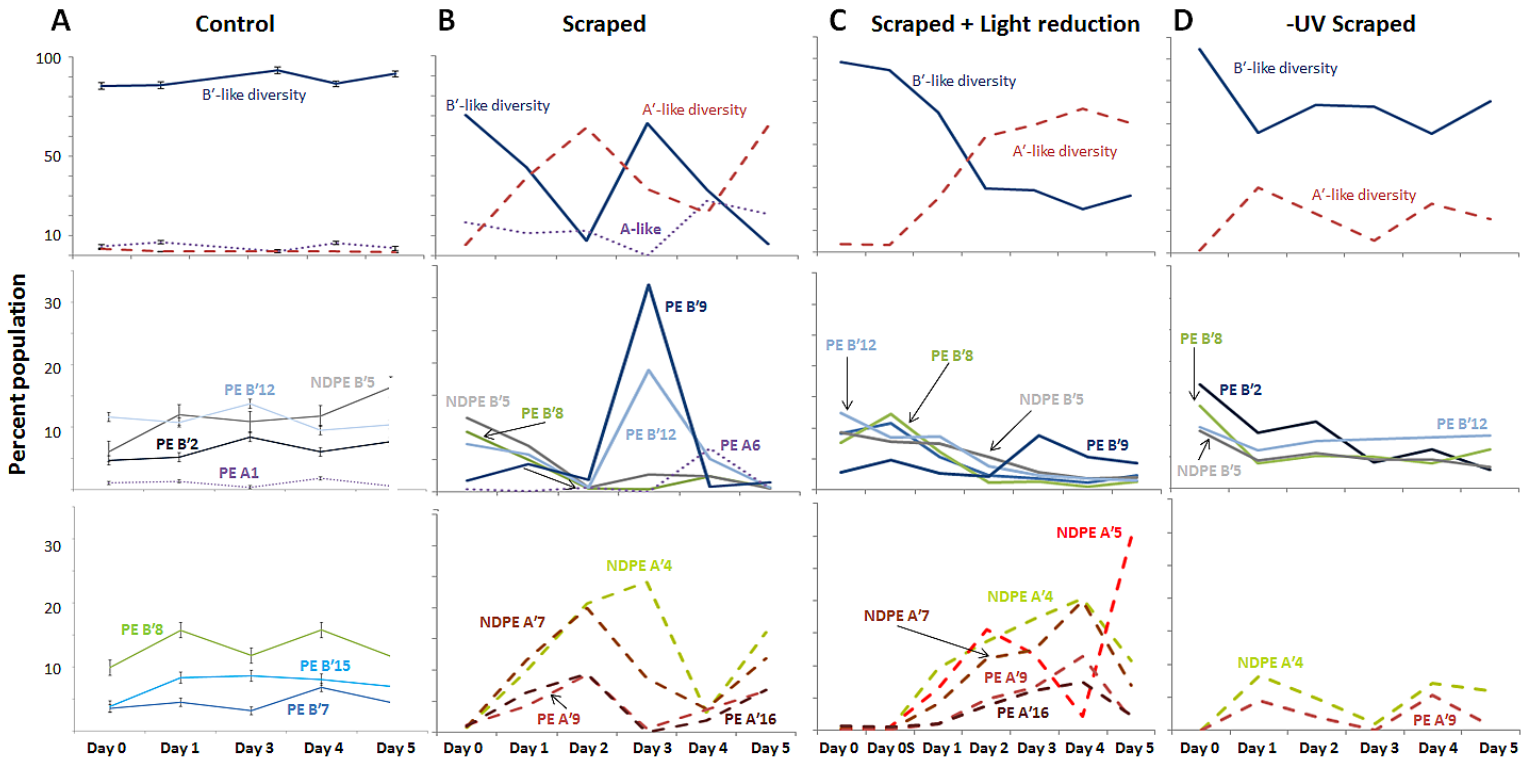


Figure 6.3. Population percentages of abundant *Synechococcus* putative ecotypes (PEs) in samples collected over a 5 day period during (A) ambient light and undisturbed (control), (B) scraped (i.e. removal of the top green layer), (C) scraped and ~92% light reduction and (D) scraped an UV exclusion conditions in a section of mat at Mushroom Spring in ecological zone (EZ) 1 (~58°C). Change in total A′, A and B′-like diversity (top); solid lines represent B′-like PEs (middle) and dashed lines represent A′-like PEs (bottom). Samples are separated into two parts for clarity (in A, middle and bottom panels report different predominant PEs; in B, C and D middle and bottom panels show B′-like and A′-like PEs, respectively). Error bars in A represent the standard error between replicate samples (n = 2).

PE B′2 compared to PE B′9, and the absence of A-like PEs in 1996 samples (see Chapter 3; Appendix B). However, the disturbance experiment was conducted on the old effluent channel in 1996 before it dried up and shifted to its current location, and PE B′9 was lower in relative abundance in the 1996 mat (Chapter 5, Figure 5.5, 1996, p.140).

Comparison of samples taken before and immediately after scraping (Figure 6.3C; compare Day 0 and Day 0S, where S indicates that the sample was collected just after scraping off the top green layer) showed similar PE relative abundances, even though the top green layer had been removed. Similar results were observed in comparing other day 0 and 1 samples (Figure 6.3B and D). This suggested that templates whose abundances had been dramatically reduced could still be amplified. A similar amplification must have occurred in the *psaA* transcription experiment described in Chapter 3, in which nighttime PE-specific transcript relative abundances resembled those detected in the evening, even though total *psaA* transcript levels were extremely low at night (Chapter 3, Figure 3.4, p.79; Liu et al., 2012).

PE Colonization and Succession
After Removal of Top Green Layer

Recovery under Ambient Light Conditions. In the unaltered control samples there was minimal change in the relative abundances of predominant PEs over the 5 days past disturbance (Figure 6.3A). B′-like diversity made up nearly all of the variation within control samples for the duration of the study. Reproducibility was reasonable in the undisturbed control samples (standard deviation was 35-53 percent of mean values; 43% average percent of means), but much poorer in samples collected after disturbance

(standard deviation was 29-89 percent of mean values; 63% average percent of means) (Figure 6.4). PE relative abundances also fluctuated significantly at some time points (Figure 6.3B-D). This irreproducibility may have been a consequence of the complex events occurring during recolonization. Low numbers of cells may have been heterogeneously dispersed across the disturbed area, adding to a background of *psaA* sequences remaining after scraping, and most predominant PEs contained multiple HFSs that contributed to the populations diversity. Additionally, competition among populations establishing themselves at different rates after initial colonization at different parts of the disturbed area may have occurred. Consistent with this hypothetical scenario, thin yellow-green to green films were noted in the initial days following disturbance, but by day 3, colonies began to appear randomly scattered throughout the site. By day 5 colonies had become so numerous that it was difficult to avoid them during collections, and in some cases patches had begun to form (Figure 6.5).

Since reproducibility was poor, mean values are shown without error bars in Figures 6.3B-D; standard deviations of abundant PEs are reported in Appendix D Tables D10 and are summarized in Figure 6.4. However, percent standard deviation between replicate samples decreased by days 4-5, indicating that initial colonization and competition may have contributed more to variation observed in the first few days after disturbance, but that populations stabilized at later time points, possibly as they began to grow.

During mat recovery after disturbance under ambient light conditions (Figure

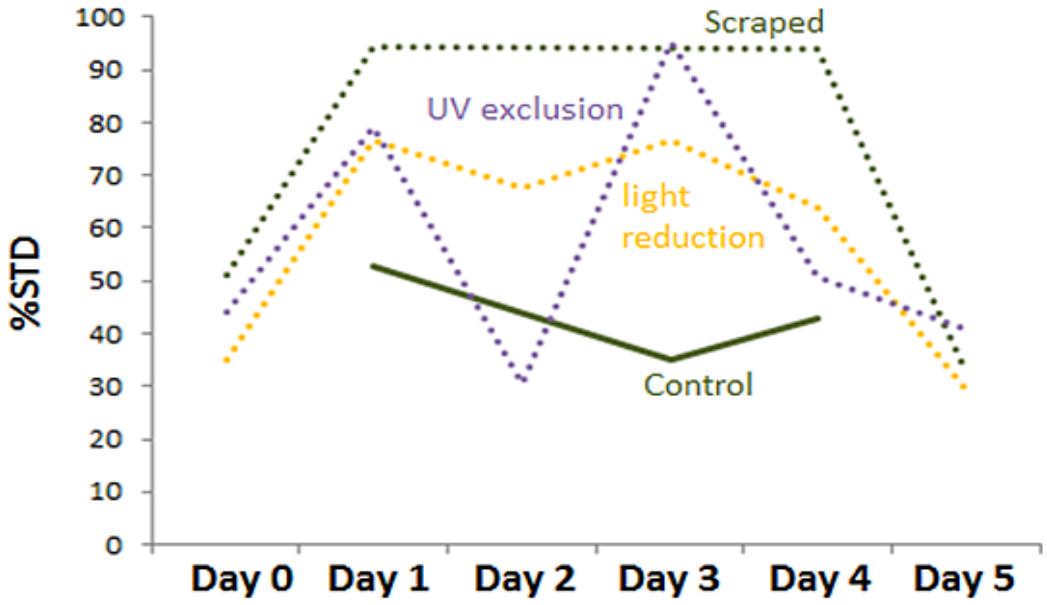


Figure 6.4. Average percent standard deviation (% STD) of all abundant putative ecotypes (PEs) for unaltered mat (solid green), scrapped ambient light (dotted green), reduced light (orange) and UV-blocked (purple) samples collected over the 5 day period. Dotted lines indicate recovery after mat disturbance (i.e. scraping off the top green mat layer under all light conditions) (also see Appendix D Table D10).

6.3B), B′-like and A′-like PEs were present in most scraped samples and fluctuated in terms of their relative abundances. A-like diversity was only observed in high relative abundance in one sample (Figure 6.3B, day 4). In the first 2 days B′-like PEs were lower in relative abundance, and A′-like PEs normally located at higher temperatures (EZ3) were higher in relative abundance at the experimental site. A′-like PEs made up ~70% of the diversity in day 5 samples (Figure 6.3B).

Despite the fact that A′-like PEs are typically found at higher temperatures, A′-like ND PE A′4, ND PE A′7 and PEs A′9 and A′16 all colonized the newly forming mat. By day 5, ND PE A′4 and ND PE A′7 were the most relatively abundant PEs in both

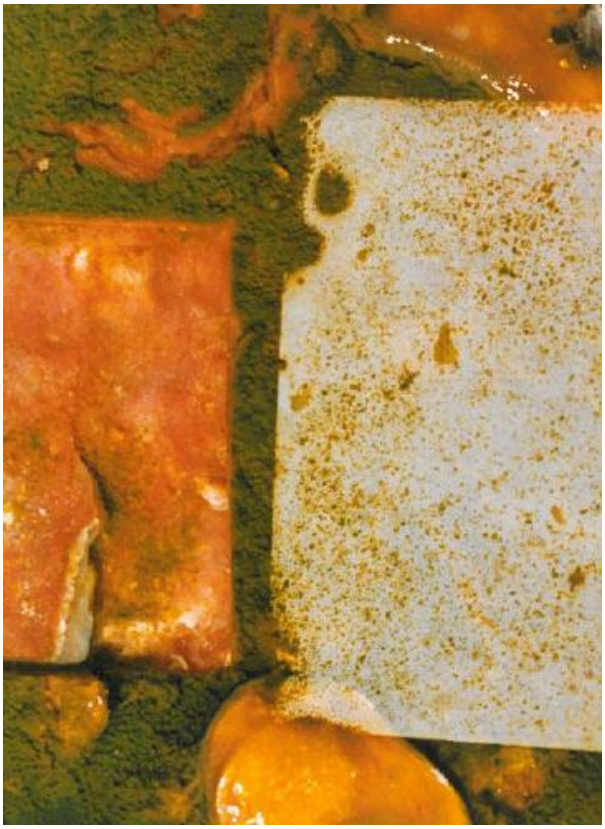


Figure 6.5. Porcelain block adjacent to scraped mat surface showing heterogeneous colony formations on both the scrapped surface and the filter several days after removal of the top green layer (Mike Ferris and David Ward, unpublished).

samples, and B′-like PEs were nearly absent. One sample contained large amounts of genetic variation at low relative abundance (e.g. Figure 6.3B), and on day 4 only PEs A6 and B′12 were ≥5% of the sample. In some post-disturbance samples, B′-like PEs that were below 5% in relative abundance before disturbance were detected in high relative abundance (e.g., PE B′9). Interestingly, although PEs usually contained multiple HFSs, a single DV was responsible for the increase in the relative abundance of PE B′9 in one replicate sample at day 3 (Figure 6.3B; Day 3). This could have been a result of clonal

growth after colonization and/or PE B′9 being a PE population of low relative abundance, with minimal HFS variation, at this site in 1996 (Chapter 5, Figure 5.5, p.140).

The majority of PEs that were capable of recolonization of newly forming mat after disturbance were PEs identified as surface populations (e.g. PEs B′8, B′9 and B′12), or higher-temperature PEs that could possibly be surface-oriented populations residing closer to the source pool (e.g. ND PE A′4, ND PE A′7 and PEs A′9 and A′16).

Recovery Under Altered Light Conditions

As shown in Figure 6.3 C and D, and comparatively in Table 6.2, similar trends were observed during recovery of disturbed mats when the light environments were altered by dimming or exclusion of UV light. In both cases, B′-like PE diversity decreased and A′-like PE diversity increased. Reproducibility was also lower than in controls (Figure 6.4). In many cases the specific PEs were similar during recovery under all light conditions. However, the differences in recolonizing PEs under different light conditions suggest that some species competed better in the different light environments. For instance, ND PE A′5 and PEs A6 and B′2 were only detected in high relative abundance during recovery under ambient, dim or UV-light removal conditions, respectively. In addition, ND PE B′5, ND PE A′7 and PEs B′9 and A′16 were detected during recovery in ambient and dim light conditions, but not when UV-light was excluded. These differences provide evidence that the recovery environment influences which species recolonize, suggesting that competition is also a factor involved in determining species composition in the recovering mat.

Table 6.2. Summary of abundant putative ecotypes (PEs) ≥5% of a sample detected during recolonization of newly forming mat after disturbance. High temperature PEs are indicated in red; known surface-oriented PEs are indicated in green. The mean percent population and percent standard deviation (%STD; in parentheses) for combined samples are reported. Arrows indicate an increase in PE relative abundance after disturbance at a time point during the duration of the study.

	Percent of population				
		1996 sites			
PE	2011 PEs in flow sites 3 and 4 (June)	1996 Control site	Scraped	~92% µE [a] scraped	-UV Scraped
B'2	3.4	6.4(45)			8.2(50)↓
B'3	3.2				
NDB'4					
NDB'5		4.5(18)	4.2(56)↓	5.2(38)↓	
B'7		4.5(57)			
B'8	2.6	13(34)	3.1(35)↓	4.4(48)↓	6.2(60)↓
B'9	3.4		7.1(25)↑	4.4(62)↑	
B'12		11(12)	6.4(40)↑	5.6(19)↓	7.9(26)
B'15		7.2(49)			
A1	6.1	1(89)			
A5	3.6				
A6			1.4(74)		
NDA'4			13(43)↑	10(41)↑	4.6(70)↑
NDA'5				9.5(83)↑	
NDA'7	13.6		9.4(48)↑	7.8(40)↑	
A'8	14				
A'9	7.5		4.3(54)↑	3.7(64)↑	2.1(83)↑
A'16	6.6		4.3(53)↑	3.1(52)↑	
%STD		35-53%	33-94	30-77	40-95

[a] µmol photons m^{-2} s^{-1}

Cells in the Flow

Water samples were analyzed to gain insight into cell dispersal at sites along the flow path. Although these analyses were made on samples collected many years after the disturbance experiment, they provide insight into patterns of dispersal of cells via water as potential colonizers of the downstream mat. A total of 14 samples (5 from OS; 9 from MS) yielded 6013 sequences (average of 430; standard deviation of 408); ES predicted a

total of 10 A′-like, 5 A-like and 7 B′-like PEs. PE diversity was lower in the water compared to corresponding mat samples (Chapter 5, Figure 5.3, p.135), and PE relative abundance in the water increased along the effluent flow path downstream from where they are typically observed in the mat (Chapter 5, Figure 5.3 and Chapter 3, Figure 3.2, pp. 135 and 77). Relatively abundant PEs in the water contained >1 HFS, which would facilitate colonization of newly forming mat by populations with diversity, as was observed for the majority of PEs in this experiment.

Mushroom Spring (June 2011). The water in the ~68-72°C source pool of MS was dominated by A′-like PEs, which made up most of the PE diversity in the water flowing above site 0 (Figure 6.6A; slanted bar fill). The most abundant A-like PEs (checkered bar fill) were found in low relative abundance at all temperature sites, possibly indicating that lower-temperature side pools contributed to the PE content of the water flow. PE A1 was the most relatively abundant A-like PE surface population in the mat at EZ2 (a region of the mat with primarily A-like PE genetic diversity located between 64 and 65°C), and was found deeper in the mat at cooler temperatures (EZ1) (Chapter 3, Figure 3.3, p.78). B′-like diversity was in very low abundance at sites 0-2, but was readily detected in water flowing above sites 3 and 4. PEs B′8 and B′9 (solid bar fill), which are relatively abundant surface-oriented B′-like populations in the mat at cooler temperatures (EZ1; Chapter 3, Figure 3.3A, pp.78), were higher in relative abundance in the water at the most downstream sites (sites 3 and 4), where the combined A′-like PE diversity totaled just ~20% of the sample. *Synechococcus* PEs in the flow at sites 3 and 4, where temperatures

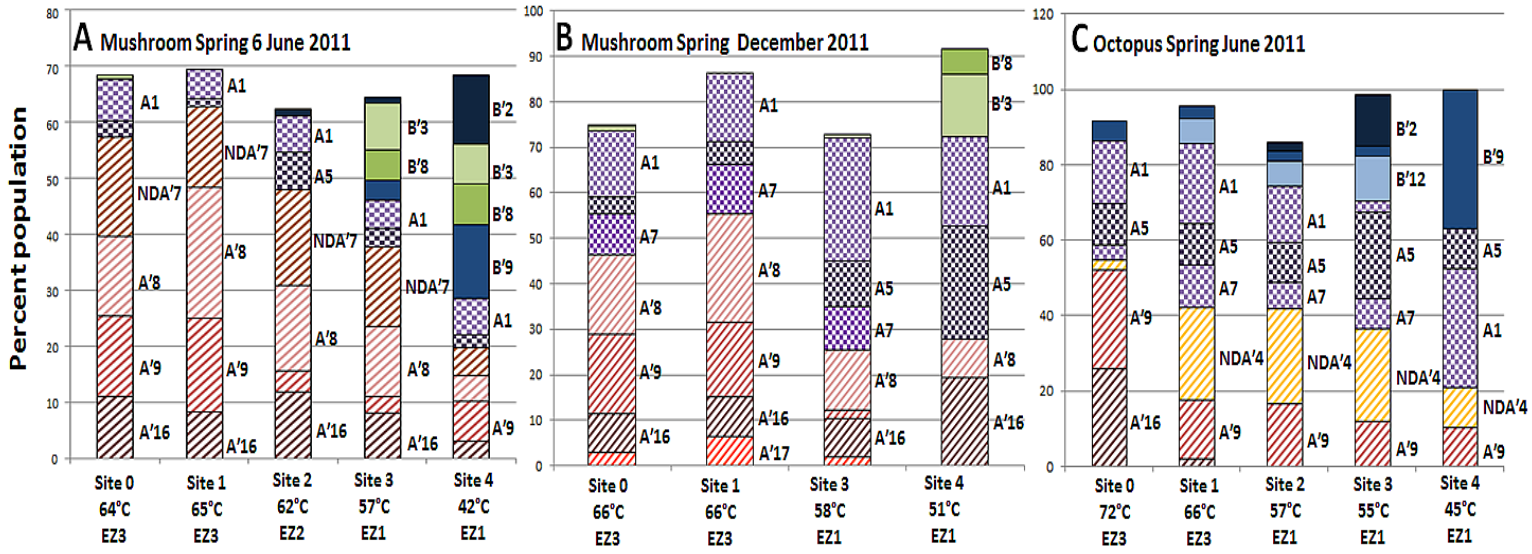


Figure 6.6. Percent population of relatively abundant putative ecotypes (PEs) in the water along the effluent flow path at Mushroom Spring on (A) 2 June and (B) 15 December 2011, and (C) at Octopus Spring on 2 June 2011. PE designations are labeled to the right of the representative bar. Missing data for site 2 in panel B result from failed sequencing reactions. Striped bars represent A′-like, dotted bars A-like, and solid bars B′-like PEs.

were comparable to that at the 1996 disturbance site, are summarized in Table 6.2 for comparison to PEs recolonizing disturbed sites in 1996. Interestingly, to the extent that 2011 samples represent PEs in the 1996 flow above this site, not all PEs in the flow colonize the disturbed mat (e.g. B′2, B′3, A1 and A5), providing additional evidence of possible ecological distinctness, as some PEs are able to colonize newly forming mat better than others.

Mushroom Spring (December). Light is reduced in the winter to ~25% of the photon irradiance observed in the summer, declining from >2000 to ~500 µmol photons m^{-2} s^{-1} average midday light intensity (see Chapter 5 Table 5.1 and Figure 5.3, pp. 135 and 135). Samples were analyzed to determine whether PE content and relative abundances along the effluent flow path in December (Figure 6.6B) were similar to those observed in June (Figure 6.6A). A′-like PEs still dominated in the water flowing above site 0, though ND PE A′7 was not detected in winter samples. All the other relatively abundant A′-like PEs were present in both December and June water samples. More A-like PEs were detected in December (e.g. PEs A1, A5 and A7), and PEs A1 and A5 had increased in relative abundance at downstream sites compared to June water samples (sites 3 and 4). While B′-like PEs were present in the lowest temperature sample (site 4), the overall B′-like PE diversity and relative abundance in the water flow was lower. PE B′9 was present at <1% of the total diversity, and the only predominant B′-like populations in the water were PEs B′3 and B′8. In December mat samples, PE A1 was observed to increase in relative abundance in EZ3, and PE A5 increased in relative abundance in EZ1 in concert with reduced B′-like diversity (Chapter 5, Figure 5.3,

p.135). Additionally, these PEs might have different vertical positioning relative to light penetration and quality at different times of the year, which could influence cellular washout and transport downstream.

Octopus Spring (June). OS water samples had similar PE distribution patterns compared to MS in June (Figure 6.6C). A′-like PEs were predominant in water flowing above site 0, though ND PE A′7 and PE A′8 were not detected in OS, and ND PE A′4 was detected in high relative abundance. A-like PEs were present in higher relative abundance in all samples analyzed compared to MS, possibly due to backwash after flow surges or to the inclusion of cells from lower-temperature side pools. Additionally, OS has larger temperature oscillations because of periodic spring overflow every 4 to 5 minutes, and this causes the upper and lower temperatures at specific locations to fluctuate ≥10°C from the average (Miller et al., 1998; Shane Nowack, data not shown). This could facilitate selection for more A-like diversity in the EZ1 mat if the upper temperature limit is the primary selective force determining ecotone boundaries and PE distributions along the effluent flow path, and could explain the higher relative abundances of A-like PEs in summer water flow at OS. B′-like PEs were relatively abundant in the lowest temperature samples, and PEs B′2, B′9 and B′12 were the most predominant populations. PEs B′12 and A7 were higher in relative abundance in OS water flow in EZ1, which was consistent with results from studies conducted over several years in OS mat samples (Chapter 5; Figure 5.6, p.144). While difficult to directly compare OS with MS, PEs B′12 and A7 were identified as low-abundance, surface-oriented populations in the MS mat (Chapter 3; Figure 3.3, p.78). PE B′9 was higher in

relative abundance further downstream at water flowing above site 4 (~45°C). PE differences between springs would possibly result from different populations being able to colonize newly forming mat after disturbance, or spring-specific adaptations to the local biotic and abiotic conditions.

Conclusions

In Ferris et al. (1997) *Synechococcus* 16S rRNA genotypes were not detected by PCR and DGGE analysis immediately after mat disturbance (Figure 6.1A). However, in this analysis *psaA* variants were never completely absent at any time point, even immediately after scraping the top green layer. It is important to point out, however, that the barcode approach provides information about relative, rather than absolute abundances of populations. The observed relative abundances of populations immediately after disturbance could have been due to the amplification and sequencing of only a few cells remaining after the disturbance, remnant DNA from these cells in deeper mat layers, or cells that are rapidly deposited onto the mat surface. As such, the relative abundances of these low-abundance populations might be more difficult to reproducibly determine.

After removal of the top green layer, some A′-like and/or B′-like PE populations were present in recolonizing mats under all light conditions. A′-like PEs were able to colonize recovering mat despite not typically being found in EZ1. Ferris et al. (1997) observed colonization of the A′ and A‴ 16S rRNA genotypes after mat disturbance. Results from Ferris et al. (1997) were verified in this experiment, as *psaA*-defined A′-like PEs colonized disturbed mats, likely because they are more abundant in the water flow.

Even though A′-like populations are not optimally adapted to temperatures corresponding to disturbed sites, it has previously been shown that they can grow at these temperatures, albeit more slowly (Miller and Castenholz, 2000).

In MS, different PEs recolonized newly forming mat under different light conditions. B′-like diversity dominated undisturbed control samples, and A′-like and A-like diversity were greater than B′-like diversity during recovery in response to scraping. A′-like diversity was higher in relative abundance, but some B′-like PEs were still competitive during recovery under reduced light conditions, and B′-like diversity was greater than A′-like and A-like diversity during recovery in the absence of UV-light. Colonization by different PEs during recovery under different environmental conditions indicates that ecological adaptation, as well as dispersal, is important during the recolonization of newly forming mat, as suggested by Morin (2011) (Figure 4.1). Some A′-like PEs that are heavily dispersed appear to have an initial advantage in the short-term, though PEs that are better adapted to the disturbed site should still have the advantage in the long-term, as A′-like PEs are typically not found in the EZ1 mat. However, to the extent that PEs in the flow in 2011 resembled those in the flow in 1996, not all heavily dispersed populations are capable of colonization, providing further evidence that adaptation is involved during initial colonization of disturbed mat.

Percent standard deviation immediately increased in response to scraping under all light conditions, though declined again by days 4-5 to match control samples. The observed differences in PE relative abundances over time between replicate samples may have resulted from sampling low numbers of cells that are deposited heterogeneously

over the recolonizing surface and/or ecological populations competing under different environmental conditions, reshaping the local community during primary succession after disturbance.

The PEs with greatest relative abundance in the water overflowing the mat at the temperatures at which the disturbance experiment was performed are the predominant high-temperature and largely surface-orientated PEs identified in previous studies (Chapter 3, Figures 3.2 and 3.3, pp. 77 and 78). The predominance of source-pool A′-like PEs and surface-oriented PEs in recovering mat samples collected at disturbed sites supports the hypothesis that highly dispersed populations that are more relatively abundant in the water flowing above the mat surface are the initial colonizers during primary succession.

The relative abundances of PEs in the water flow in December compared to June are similar to the results presented in the seasonal study of PEs in the mat (Chapter 5, Figure 5.3, p.135), further indicating that light is a powerful selective force in determining PE abundances and distributions seasonally. As a result, PE population distributions differ between summer and winter months along the effluent flow path, and possibly along vertical gradients as well, which would have an effect on dispersal and initial colonization of damaged mat after disturbance in different seasons. However, more studies are needed to analyze the vertical distributions and population dynamics of *Synechococcus* ecotypes in different seasons.

OS and MS have similar chemistries but different hydrodynamic properties, which could influence PE distributions along the effluent flow path. Additionally, while

PE diversity in the mat is shared between springs (Chapter 5, Figures 5.5 and 5.6, pp. 140 and 144), predominant PEs have different relative abundances at similar sites defined by average temperature, which could be due to spring-specific PE adaptations acting on these populations independently. Differently adapted PEs are at varying relative abundances along the effluent flow paths, and the different hydrodynamic properties between these springs may mix PEs from different EZs, such that both factors probably contributed to the observed differences in PE relative abundance in the water flow (Figure 6.6A and C). These results could indicate that different PEs are more or less likely to be dispersed in the water flow and to recolonize newly forming mat at downstream sites in OS and MS after environmental disturbance.

CHAPTER 7

COLLABORATIVE TI454-BARCODING PROJECTS

During my dissertation research I was also involved in three additional projects involving barcode applications, which are described in this chapter: (i) assessing whether or not *Synechococcus* cultures were unicyanobacterial before genome sequencing, (ii) analyzing distributions of genetic loci used in previous population genetics and multi-locus sequence analysis (MLSA) studies of *Synechococcus* populations (Melendrez et al., 2011; Melendrez et al., in prep) and (iii) using *Synechococcus* genome sequences to link distributions of MLSA and *psaA* loci.

Synechococcus Culture Work and Obtaining Ecotype-specific Genomes

Rationale

In order to test hypotheses about ecotype-specific adaptations infered from environmental distribution studies, and to understand the underlying genomic basis for these adaptations, individual *Synechococcus* isolates corresponding to ecological populations predicted from the *psaA* locus were obtained using filter-cultivation methods (de Bruyn et al., 1990) (Figure 7.1), and their genomes were sequenced with support from a Department of Energy Joint Genome Institute Community Sequencing Program award. The proposal was to obtain complete genomes of 18 cultivated isolates representative of 3 strains for each of 6 putative ecotypes (PEs; populations predicted through evolutionary

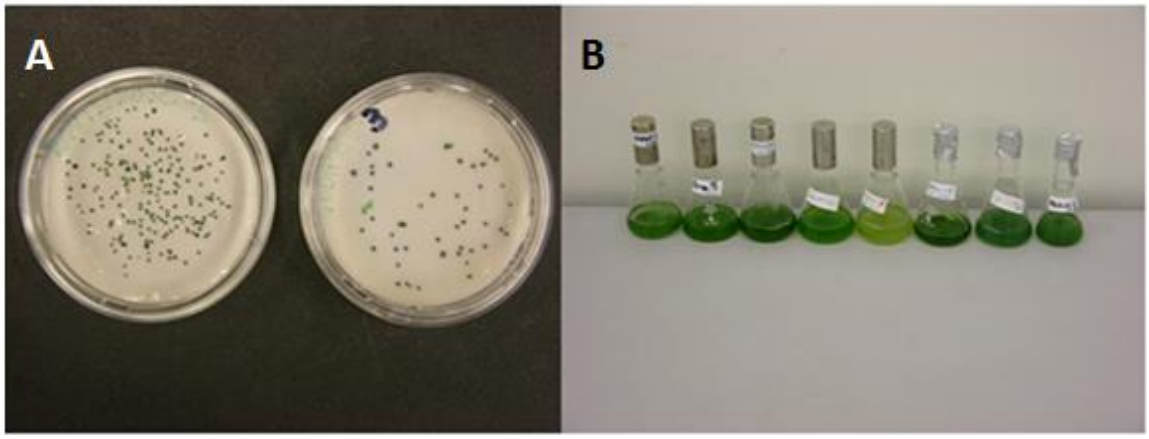


Figure 7.1. *Synechococcus* cultures. (A) Separated colonies on 47 mm polycarbonate filters. Left and right plates are from 10^{-3}- and 10^{-4}-fold mat dilutions, respectively. (B) 50 ml Erlenmeyer flasks (20 ml medium) of *Synechococcus* cultures that were grown at 52°C and at ~35 to 50 µmol photons m^{-2} s^{-1} exhibiting phenotypic differences.

simulation analysis to be ecologically distinct) in order to permit comparative analyses of replicate strains in different species. It was of utmost importance to obtain unicyanobacterial cultures in this work. Genomes that had been obtained previously were from mat *Synechococcus* isolates that had been cultivated from low-dilution mat samples (i.e., samples containing 10^{10} *Synechococcus* cells/ml that were only diluted by a factor of 10^{-2} to 10^{-3}) (Allewalt et al., 2006; Bhaya et al., 2007), and it was suspected that these genomes might be based on multiple ecotypes. In order to test whether these and new isolates were unicyanobacterial, Ti454-barcode high-throughput sequencing was used to obtain deep sequence coverage at the *psaA* locus to examine whether or not the genetic variation in a culture was consistent with expectations for a strain of a single ecotype (see below). A pipeline to obtain unicyanobacterial cultures for genome sequencing was established (Figure 7.2), and a team was formed consisting of current doctoral student

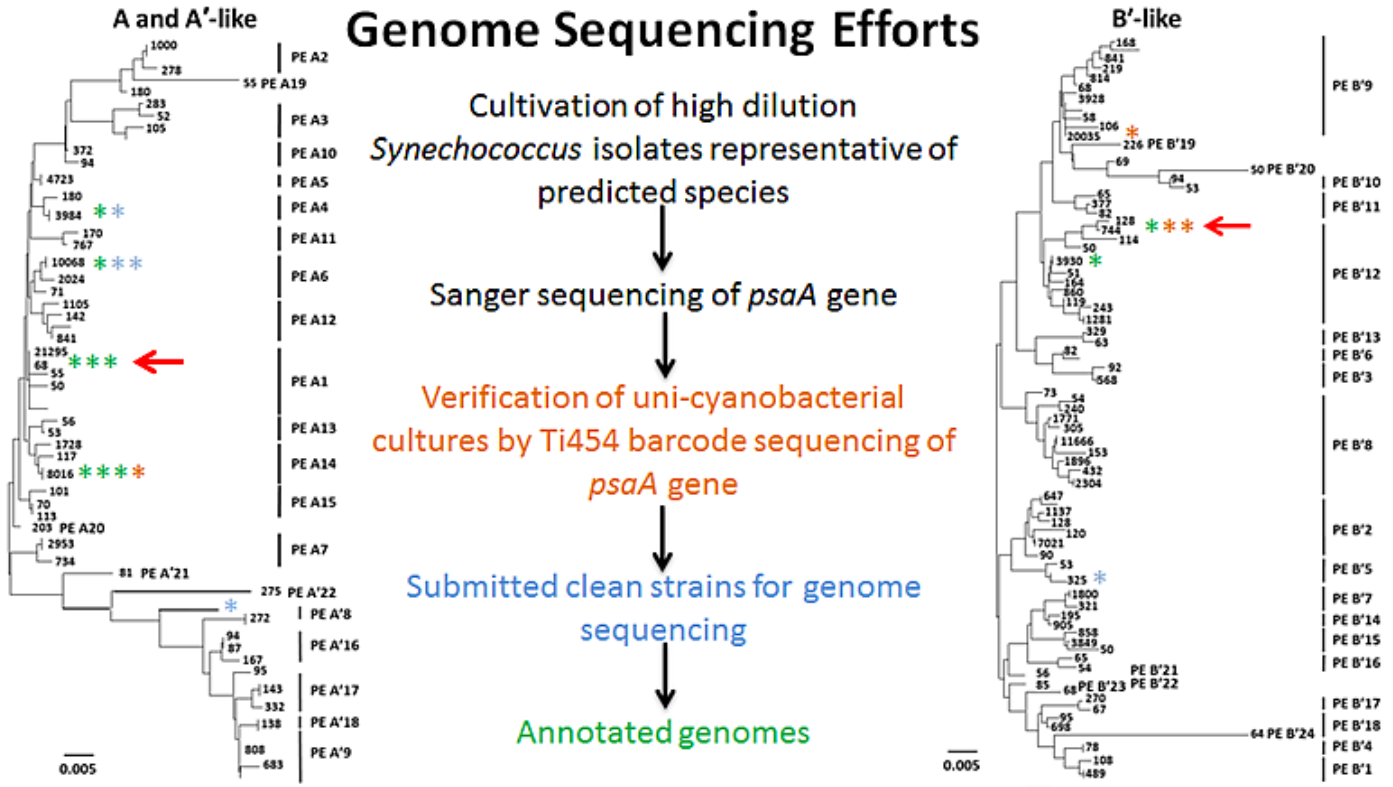


Figure 7.2. Neighbor-joining phylogenies of *psaA* A-like and A′-like (left), and B′-like (right) high-frequency sequences (≥50 copies in all combined samples) and corresponding genome sequences. Putative ecotypes (PEs) are demarcated with vertical bars. Colored asterisks next to PE dominant variants correspond to where in the purification/genome sequencing pipeline the culture currently resides (center). Red arrows indicate the B′ genome JA-2-3B′a (2-13), and the A genome JA-3-3Aa. The A genome is identical to the dominant variant sequence of PE A1.

Shane Nowack (sample dilution, isolation and culturing), current master's student Millie Olsen (molecular biology and comparitive genomics) and myself (molecular biology and Ti454-barcode sequence analysis) to achieve this goal.

Materials and Methods

Samples for cultivation were collected from Mushroom Spring (MS) on 10 September 2008, 7 September 2011 and 2 October 2012. *Synechococcus* strains (Table 7.1) representative of different PEs detected in earlier *psaA* studies (Figure 7.2; also see Chapter 3, Figure 3.1, p.75) were cultivated and purified by Shane Nowack.

Cultures were first analyzed at the *psaA* locus using Sanger sequencing to determine the predominant cyanobacterial sequence in each culture. In this case, DNA was extracted and the *psaA* locus was PCR-amplified as described in Chapter 2, and amplicons were sent to the Idaho State University Molecular Biosciences Core Facility for Sanger sequencing. Sequence purity was estimated by analyzing the sequence chromatograms and associated nucleotide quality scores. Culture sequences that had single chromatogram peaks at all nucleotide sites, and high nucleotide quality scores, were further analyzed by Ti454-barcoding and sequencing at the Research and Testing Laboratory, Lubbock, TX, as described in Chapter 4, where unique barcode sequences were ligated onto the 'psaAcenterforward' primer (Chapter 3 Material and Methods) before PCR amplification for each culture, and amplicons were sequenced. This yielded an average of 2,628 sequences per sample (ranging from 449 to 7,669 sequences). Using ClustalW to generate alignments, sequences were processed through an error-fixing Perl script (Homopolymer Extinguisher) as described in Chapter 3. Sequences were aligned

and manually edited using Sequencher 4.7, and Neighbor-joining trees were constructed using MEGA 5.0 software as described in Chapter 4 Materials and Methods. Cultures shown to be unicyanobacterial were sent to the Joint Genome Institute for genome sequencing.

Table 7.1. Summary of *psaA* Ti454-barcode and sequencing analyses of *Synechococcus* cultures. Sequences with systematic errors were excluded.

		dominant variant sequences [b]			other sequences	
Strain name [a]	**total sequences**	**Putative Ecotype**	**DVs**	**%**	**total**	**%**
JA-2-3B′a (2-13)	2054	all	1690	82.3	364	17.8
		€	1385	67.4		
		B′ 24	300	14.6		
		B′11	5	0.2		
JA-3-3Aa	1116	all	931	83.4	185	16.5
		A1	930	83.3		
		A14	1	0.1		
65AY6A5	1325	all	1106	83.5	219	16.5
		A4	1105	83.4		
		A1	1	0.1		
63AY4M2	2027	A6	1756	86.6	271	13.4
63AY4M1	1007	A14	816	81.0	191	19.0
60AY4M2	443	A14	374	84.4	69	15.6
65AY640	920	A14	756	82.2	164	17.8
60AY6Li	980	A1	780	79.6	200	20.4
63AY4Ap-74	2244	A1	1977	88.1	267	11.9
68DH1S1	843	B′12	703	83.4	140	16.6

[a] Strain JA-3-3Aa corresponds to the strain A genome, and JA-2-3B′a (2-13) corresponds to the strain B′ genome.

[b] Sequences that are identical to the expected dominant variant (DV; identical sequence that made up the plurality of a putative ecotype (PE) clade).

€ A dominant variant sequence detected in culture that has not been identified.

Ten cultures were initially selected for Ti454-barcode analysis, including strains JA-3-3Aa and JA-2-3B′a (2-13) from which the two previously sequenced *Synechococcus* genomes (Bhaya et al., 2007) had been obtained. A culture was

considered to be unicyanobacterial if it contained a single dominant variant (DV; identical sequence constituting the plurality of a PE clade) at the *psaA* locus that is representative of a PE found in nature (x axis in Figure 7.3), as well as closely related less-abundant genetic variation (1 or 2 randomly distributed nucleotide substitutions that formed a clade with the DV). This is shown by green dots in Figure 7.3, where alignments for random selections of non-identical sequences of two different strains are shown on individual blue lines. It is unlikely that all observed diversity arose from natural mutations that occurred during growth in culture because they occur much more frequency than expected given natural mutation rates (Drake et al., 1998); PCR and sequencing error probably contributed significantly to the sequence variation. However, single nucleotide variation DV in a culture that was the result of mutation or sequencing error does not erode confidence in the determination that a culture was unicyanobacterial. In addition, all Ti454-barcode sequence data contained systematic sequencing errors at error-prone nucleotide positions. This is shown by red dots in Figure 7.3, which indicate nucleotide substitutions, or a combination of polymorphisms, that were contained in many sequences of all isolate strains at specific error-prone sites in the amplified region of the *psaA* locus. Note that the strains shown in Figure 7.3A and B have polymorphisms at the same nucleotide positions. The frequency of systematic errors was conserved across all isolate cultures sequenced, and the most common error in one isolate culture was also the most common error in all isolate cultures. This artificial variation was disregarded because was it not representative of genetic variation found naturally in the culture.

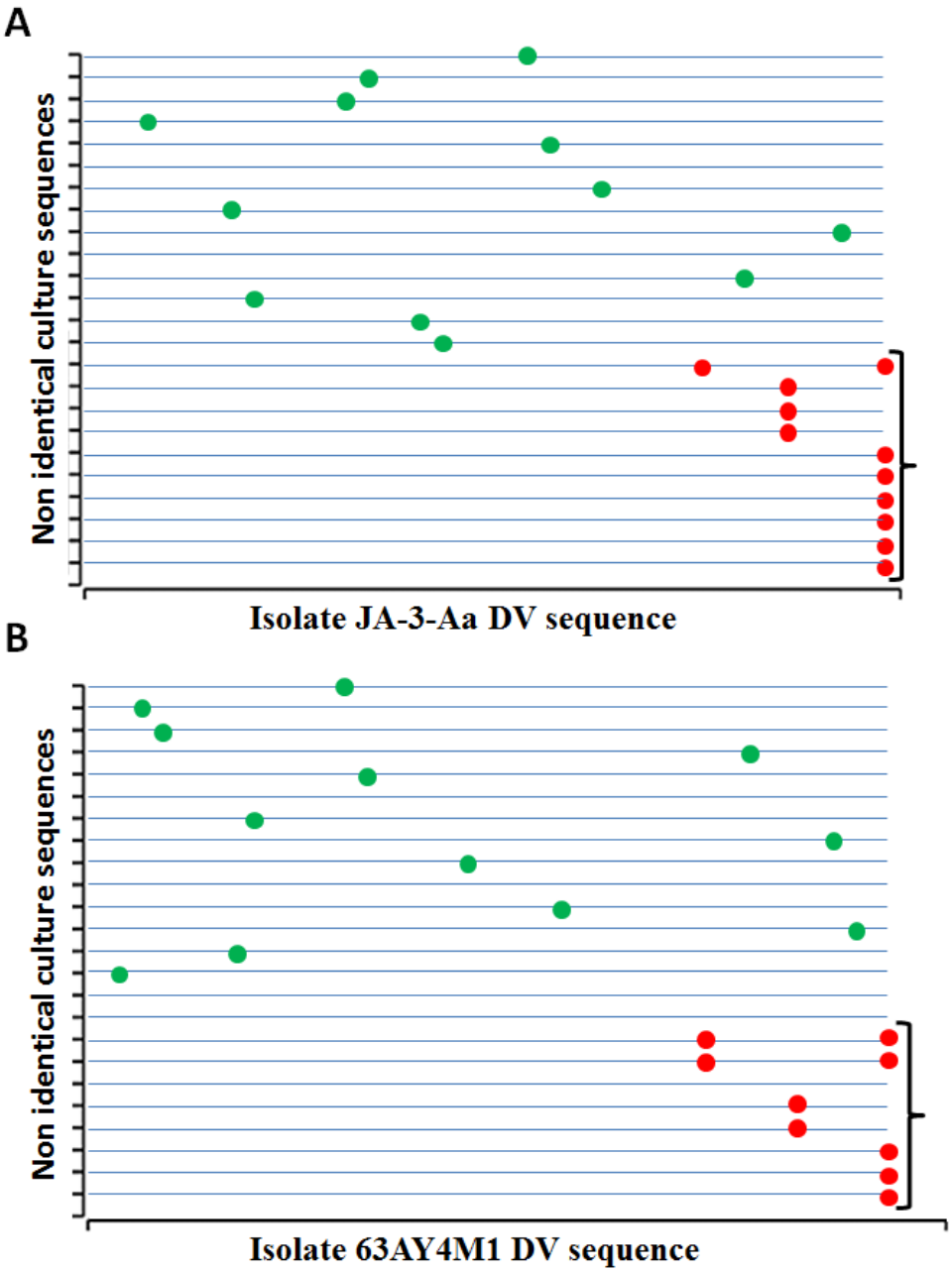


Figure 7.3. Alignments (positions 512-814; total gene length 2268 bp) of random subsets of non-identical *psaA* sequence variants (individual blue lines) generated from Ti454-barcode sequencing of isolate cultures (A) JA-3-3Aa and (B) 63AY4M1 relative to the *psaA* dominant variant (DV) sequence of each strain (base line). Dots indicate single-nucleotide polymorphisms; green dots indicate random nucleotide substitutions or sequence errors; red dots within brackets indicate systematic errors observed in the same positions in relation to the DV sequences of both strains.

Results and Discussion

The *Synechococcus* sp. strain B′ isolate (JA-2-3B′a (2-13)) contained 3 phylogenetically distant DVs, two of which corresponded to the DVs of different PEs (Table 7.1). The majority of sequences were of a sequence type that had not been previously detected, which was a unique sequence-type that clustered with PE B′12 (Figure 7.2, red arrow); the culture also contained a large number of sequences that matched the DVs of PEs B′11 and B′24, indicating that the culture was not unicyanobacterial (Table 7.1, p.182; also see Chapter 3, Figure 3.1, p.75). In contrast, the *Synechococcus* sp. strain A isolate (JA-3-3Aa) exhibited a single abundant sequence-type that was identical to the DV of PE A1. Out of 1,552 total sequences a single DV sequence representative of PE A14 was present (0.06%), possibly a result of cross-contamination. This gave me confidence that the A genome was predominantly unicyanobacterial and representative of 1 PE, but raised questions about the B′ isolate genome, which may be an amalgam of multiple populations from the environment.

To date, eight new isolates have been identified that meet the criteria established for whole-genome sequencing (Table 7.1). Cultures representative of seven closely related A-like PEs and one B′-like PE were sequenced, assembled and annotated yielding near-complete ecotype-specific genomes. Preliminary genome comparisons done by Millie Olsen have shown that strains representative of some PEs have similar gene content, including genes unique to that PE. Such genomic differences allow us to form hypotheses about what differentiates ecotypes. Strain-specific differences have also been observed, which might cast doubt on some hypothesized PEs being true ecotypes (see

below). In addition, phenotypic studies done by Shane Nowack are being conducted to observe differences in the growth rates of strains of the same and different PEs in response to varying temperature and light conditions. Preliminary studies have shown that strains of closely related A-like PEs grow optimally at different light intensities that correspond to their *in situ* vertical distributions (see Chapter 3, Figure 3.3, p.78).

Conclusions

The development of a pipeline to obtain ecotype-specific genomes was largely dependent upon the utilization of Ti454-barcode technology to determine which cultures were unicyanobacterial (Figures 7.2 and 7.3). Ecotype-specific genomes can be used to analyze genomic properties, such as gene content and arrangement, and to hypothesize ecotype-specific adaptations. Also, isolates bearing these genomes can be grown under differing environmental conditions *in vitro* (e.g. temperature and light) to observe whether strains of an ecotype have phenotypic properties consistent with their *in situ* distributions.

As a warning to future researchers, while Ti454-barcoding and sequencing revealed that the predominant *psaA* sequence variant in the culture from which the *Synechococcus* sp. strain A genome was obtained matched the DV of PE A1, the culture from which the *Synechococcus* sp. strain B′ genome was derived consisted of DVs from multiple PE found in the environment, indicating that the B′ genome is likely an amalgam of more than one population.

Distributions and Population Dynamics of Single Loci Used in Multi-Locus Analyses

Rationale

Former Ward Lab doctoral student Melanie Melendrez used bacterial artificial chromosomes (BACs) to clone large genomic inserts ranging from ~90-120 kb, which allowed access to multiple genes from the same individual *Synechococcus* cells inhabiting the mat. This, in turn, permitted MLSA of concatenated loci, which increased molecular resolution because the overall length of the sequence analyzed was greater. It also buffered ecological species demarcation against the effects of recombination by dampening the effect of recombinant genes causing individuals within a PE to be displaced from their respective clade in a phylogeny (Melendrez et al., in prep). Multiple loci located on the BAC clones were PCR amplified and sequenced, and the sequences were concatenated and analyzed by Ecotype Simulation (ES; an evolutionary simulation program based on the Stable Ecotype Model of species and speciation described in Chapter 1).

Melendrez et al. (in prep.) found that despite being an important factor influencing the evolutionary history of mat *Synechococcus* populations, genetic recombination did not erode the ecological distinction of populations predicted to be different ecotypes by ES analysis. However, the study was limited to 71 BAC clones collected from only two habitats (60 and 65°C), severely limiting the analysis of ecological distinctions to differentiate ecotypes. To enhance the Melendrez et al. (in prep) study, Ti454-barcoding and sequencing of single-loci used in MLSA analyses was used

to analyze the distributions and population dynamics of PEs predicted from MLSA loci in different temperature-defined ecological zones (EZs), which I termed EZ1 between ≤60 and 63°C, EZ2 between 64 and 65°C and EZ3 at ≥66°C, separated by two ecotones (located between 63-64°C, and 65-66°C; Chapter 3, Figure 3.2, p.77). Ti454-barcoding increased the number of habitats sampled and the sequencing depth across all habitats, and allowed me to test hypotheses about *Synechococcus* species related to differences between MLSA and single-locus analyses (SLA).

Materials and Methods

PCR primers described by Melendrez et al. (2011) were used to amplify 4 loci (*rbsK*, *aroA*, *pcrA*, and a conserved hypothetical protein locus (*chp* [locus tag CYA_2291]) from DNA extracted from samples used to study distribution of *psaA* variation relative to flow and vertical gradients (Chapter 3, Figures 3.2 and 3.3, pp. 77 and 78), and temporally in response to ~92% ambient light reduction (Chapter 3, Figure 3.5, p. 82) (83 total samples). All loci were sequenced using Ti454-barcoding technology as described in Chapter 3 Materials and Methods. The attempt to sequence multiple loci with multiple primers during a single sequencing run (i.e. multiplexing) caused many failed sequencing reactions, possibly because of differences in the energetics of primer binding (Appendix E, Section I; Appendix E Table E1). Failed sequencing reactions due to multiplex sequencing limited the study to the *rbsK* and *chp* loci.

chp sequences were obtained only from samples collected along the MS effluent flow path at 1°C intervals from 60 to 65°C in 2006 (identical samples in Chapter 2, Figure 2.3, p.47). *rbsK* MLSA-locus sequences from Melendrez et al., (in prep.) were

trimmed from 534 to 362 bp at the 5′ and 3′ ends to match Ti454-barcode sequence average length, and this caused a slight loss in molecular resolution, collapsing BAC-derived *rbsK* SLA PEs 1, 6 and 7 (Figure 7.4B) into barcode-derived *rbsK* HFS PE1 (Figure 7.4C). Barcode sequences were cleaned, and high-frequency sequences (HFSs; ≥50 copies in all combined samples) were identified. PEs were demarcated, and all HFSs and associated low-frequency sequences (LFSs; <50 copies in all combined samples) contained within a PE were summed, and population percentages were calculated as described in Chapter 3 Materials and Methods. All PE distributions shown in figures are color-coded to match MLSA PE designations shown in Figure 7.4A.

Results and Discussion

After cleaning, there were 65,372 *rbsK* sequences, yielding 88 HFSs identified from 52 samples. Samples were from vertically dissected mats located at three different average midday temperature sites (60, 63 and 65°C), and from collections made over a 6 day period after ~92% ambient light reduction.

Comparison of BAC-derived and Barcode-derived *rbsK* Phylogenies. The Ti454-barcode-derived *rbsK* A-like HFS phylogeny (Figure 7.4C) generally matched the structure of the BAC-clone-derived *rbsK* SLA phylogeny (Figure 7.4B), with the former yielding 11 PEs (with 3 singleton HFSs) and the latter yielding 7 PEs. Newly demarcated *rbsK* PEs 8 and 9 were possibly detected due to increased sequencing depth and habitat coverage. Alternatively, the presence of *rbsK* PEs 8 and 9 could be due to temporal differences in population relative abundances (see Chapter 5), and the possibility that these PEs were higher in relative abundance in

September 2008 barcode studies compared to samples that were collected in October 2003 for MLSA. The presence of additional *rbsK* PEs could also be due to the shift in the main effluent flow path at MS, which likely had an influence on the population dynamics of PEs in the system over the years studied (Chapter 5).

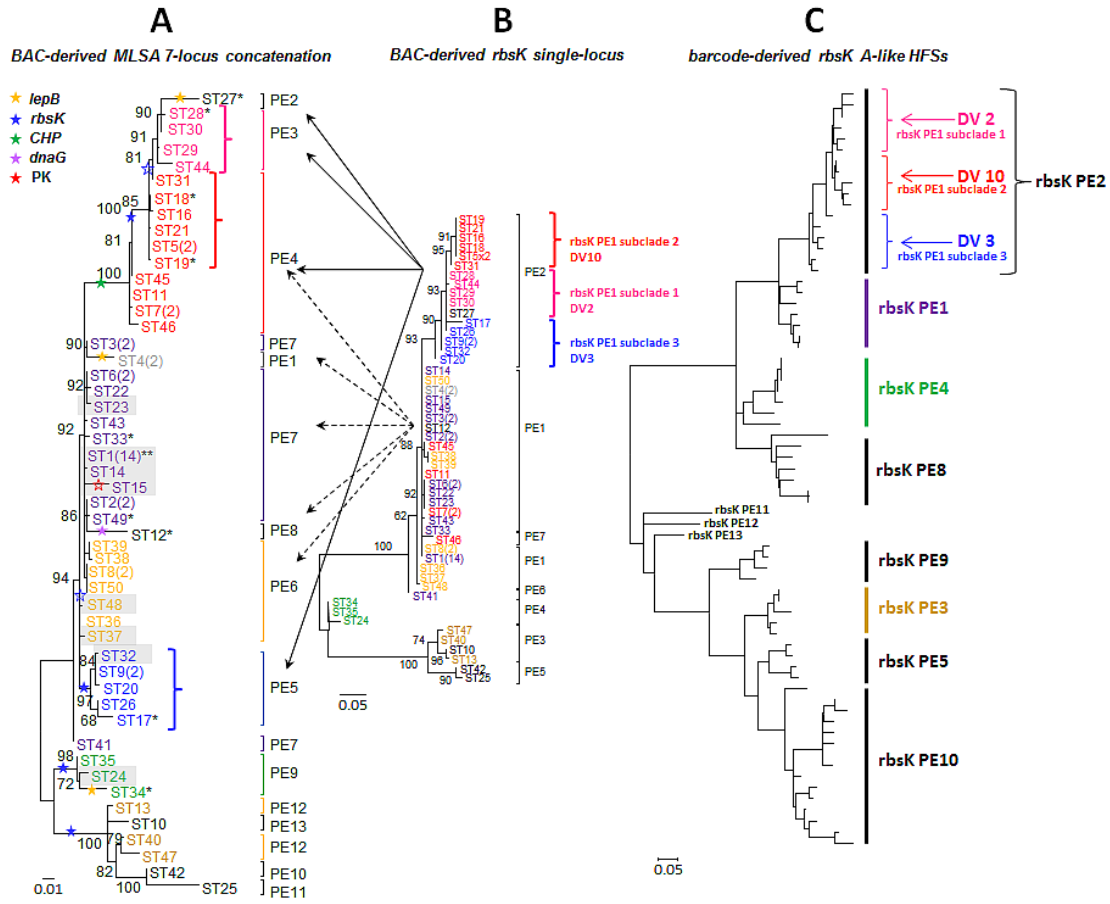


Figure 7.4. Comparison of putative ecotypes (PEs) demarcated from the (A) bacterial artificial chromosome (BAC)-derived multi-locus sequence analysis (MLSA) phylogeny, (B) BAC-derived single-locus analysis (SLA) *rbsK* phylogeny and the (C) barcode-derived *rbsK* high-frequency sequence (HFS) SLA phylogeny. Arrows match sequence type (ST) distribution between BAC-derived MLSA and SLA phylogenies. Smaller brackets embedded within *rbsK* SLA PE A2 (which corresponds to barcode-derived *rbsK* HFS PE A2) indicate subclades that were separated into distinct MLSA PEs. Scale bars correspond to nucleotide substitutions per site indicated. (A and B are modified from Melendrez et al, in prep).

ES analysis of concatenated MLSA sequences separately demarcated MLSA PEs 3, 4 and 5 (red, pink and blue variants in Figure 7.4A), which were subclades embedded within *rbsK* PE A2 in both the BAC-derived and barcode-derived HFS single-locus analyses (Figures 7.4B and C). Since the DVs of these *rbsK* PE A2 subclades correspond to different MLSA PEs, I used Ti454-barcode distribution and population dynamics analyses to test whether *rbsK* sequences unique to each of these MLSA PEs can be observed to be ecologically distinct.

Distributions of *rbsK*-defined PE A2 Subclade Populations. Analyses of vertical distributions of barcode-derived *rbsK* PE A2 subclade populations in samples collected at 60, 63 and 65°C showed that the subclades corresponding to MLSA PEs 3, 4 and 5 had unique temperature and vertical distributions (Figure 7.5). MLSA PE 4 maximizes near the upper regions of the mat, and is in greater relative abundance at 63°C (EZ1). MLSA PEs 3 and 5 both appear to maximize in the middle and lower mat regions, with MLSA PE A3 being in high relative abundance only at 60°C (EZ1). MLSA PE 5 was in high relative abundance at temperatures above and below the 63-64°C ecotone boundary. Temperature and vertical distributions of *rbsK* variants of different subclades within *rbsk* SLA PE A2 provided evidence that the associated MLSA PEs were indeed indicative of ecologically distinct populations.

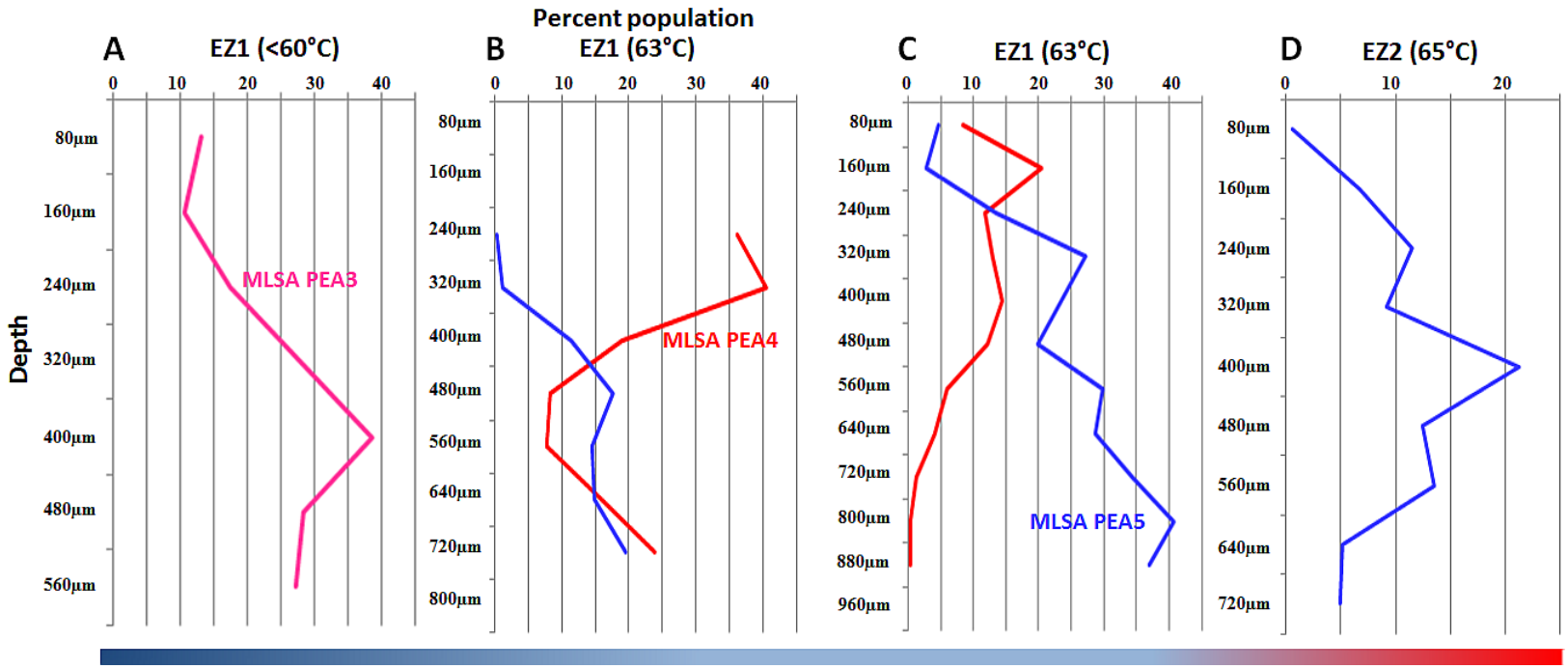


Figure 7.5. Vertical distributions at (A) 60°C, (B and C) 63°C and (D) 65°C sites in Mushroom Spring mat of combined *rbsK* variants (high-frequency sequences + low-frequency sequences) that were ≥5% in a sample from barcode-derived *rbsK* putative ecotype (PE) A2 subclades that correspond to multi-locus sequence analysis (MLSA) PEs A3, A4 and A5. Colors correspond to the same PE in all parts of the figure.

Population Dynamics of
rbsK-defined PE A2 Subclades in
Response to ~92% Light Reduction. In response to ~92% light reduction, *rbsK* variants corresponding to MLSA PE A5 decreased in relative abundance, while variants corresponding to MLSA PE A3 increased in relative abundance between days 0 and 4 (Figure 7.6). Condensation on the light altering screens may have caused light scattering, which could have also influenced changes in PE relative abundances during the sampling period. All PE distributions and responses to environmental perturbations are summarized in Table 7.2.

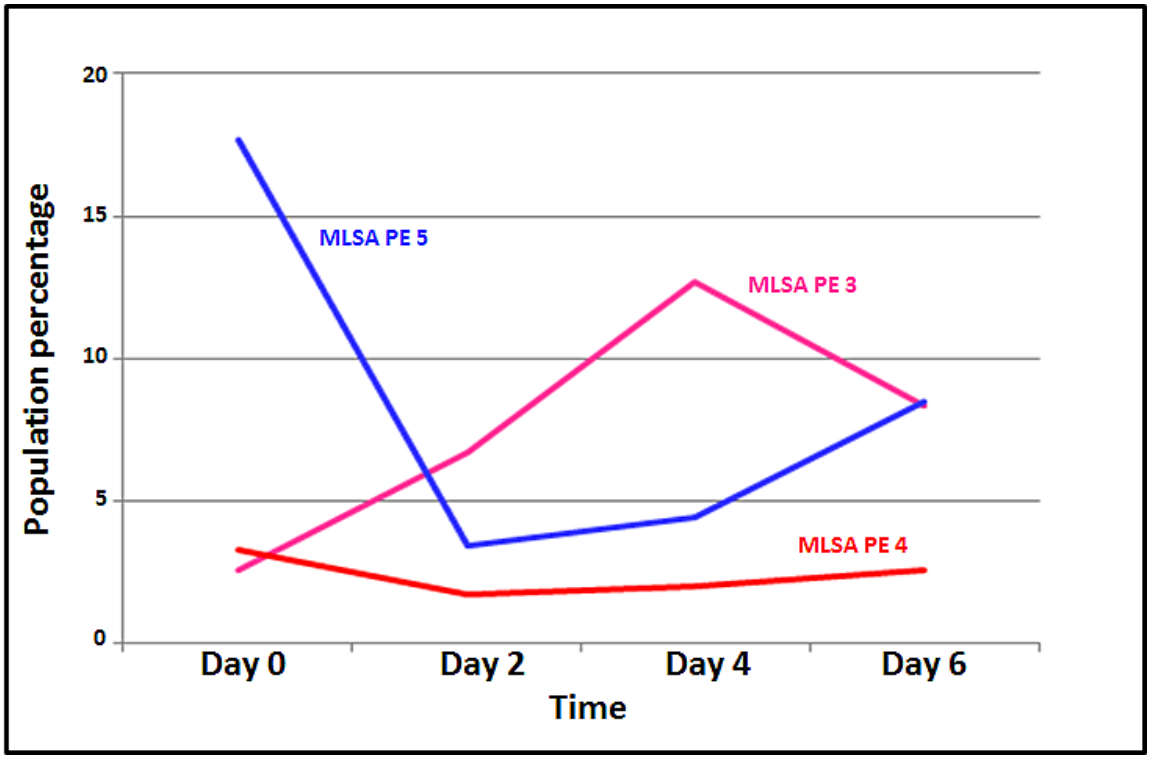


Figure 7.6. Changes in relative abundances in response to ~92% light reduction of bacterial artificial chromosome (BAC)-derived multi-locus sequence analysis (MLSA) putative ecotypes (PEs) A3, A4 and A5, corresponding to BAC and barcode-derived *rbsK* single-locus PE 2 subclades 2, 1 and 3, respectively. Study was conducted at Mushroom Spring (MS) in ecological zone (EZ) 2 (also see Chapter 3, Figure 3.5, p.82).

Table 7.2. Summary of *psaA* and multi-locus sequences analysis putative ecotypes (PEs) (using *rbsK* high-frequency sequences + low-frequency sequences) distributions and population dynamics in response to ~92% light reduction.

***psaA* PE**	***rbsK* PE**	**MLSA PE**	**60°C *psaA***	**V** [a]	**63°C *psaA***	**V**	**65°C *psaA***	**V**	**68°C *psaA***	**60°C rbsK**	**V**	**63°C *rbsK***	**V**	**65°C *rbsK***	**V**	**92% light reduction *psaA***	**92% Light reduction *rbsK***
A1	PEA3,8,11	PE13	2.6	↓	2.9	↓	26.8	↕	4.7								
A4	PEA1	PEA7	2.7	↓	0.5		3.6	↓	0.6	7.1		13.0	↓	6.0	↓	↓	↓
A6	PEA2	PEA5	6.4	↓	3.2	↓	13.2	←	1.9	22.8		10.0	↑	16.0	↑	↓	↓
A14	PEA4	PEA9	2.0	↓	1.3	↓	10.7	←	0.6	3.3		8.5	↓	10.0	↓	↑	↑
	PEA1	PEA3								13.0	↓	2.5		1.5			↑
	PEA1	PEA4								2.8	↑	6.1	↑	2.7	↑		↔

[a] "V" indicates vertical positioning at temperature specified in column to the left. Arrows under 'V' indicate high relative abundance in upper mat layers (↑), high relative abundance in lower mat layers (↓), high relative abundance in center layers (↔), and even relative abundance through all layers (↕). Perturbation responses are indicated by ↑ for increase or ↓ for decrease in relative abundance.

Comparison of BAC-derived and Barcode-derived *chp* Phylogenies. A conserved hypothetical protein-encoding locus *chp* was also analyzed along the flow path at 1°C intervals from 60 to 65°C. ES analysis of BAC-derived *chp* sequences predicted 2 PEs (Figure 7.7B), one of which was associated with MLSA PEs 2, 3 and 4, and the other of which contains variants from all other MLSA PEs (compare Figures 7.7A and B; in regions indicated by the arrow). Ti454-barcoding HFS analysis of the *chp* locus yielded 4 PEs (including two singleton PEs), and had a similar phylogenetic structure to the SLA phylogeny (compare Figures 7.7B and C). More HFS variation was detected around the PE DVs in the barcode-derived HFS phylogeny, which was probably the result of deeper sequencing.

Distribution of *chp*-defined PEs along the Flow Path. The most predominant *chp*-defined PEs A1 and A2 were distributed similarly at 60-63°C (EZ1), but differently at 64-65°C (EZ2) (Figure 7.8). *chp* PE A2 (containing MLSA PEs 2, 3 and 4) relative abundance was only ~10% of the sample at and above the 63-64°C ecotone boundary, though was ~40% of the sample at lower temperatures (EZ1). Alternatively, *chp* PE1 (containing MLSA PE5) totaled ~90% of the sample in EZ2 (64 and 65°C), but also comprised ~60% of samples collected in EZ1 (60-63°C).

Conclusions

Barcode-derived *rbsK* sequences corresponding to MLSA PEs showed unique distributions and responses to environmental perturbation, providing evidence that the subclades embedded within *rbsK* SLA PE A2 that were separated by the more highly

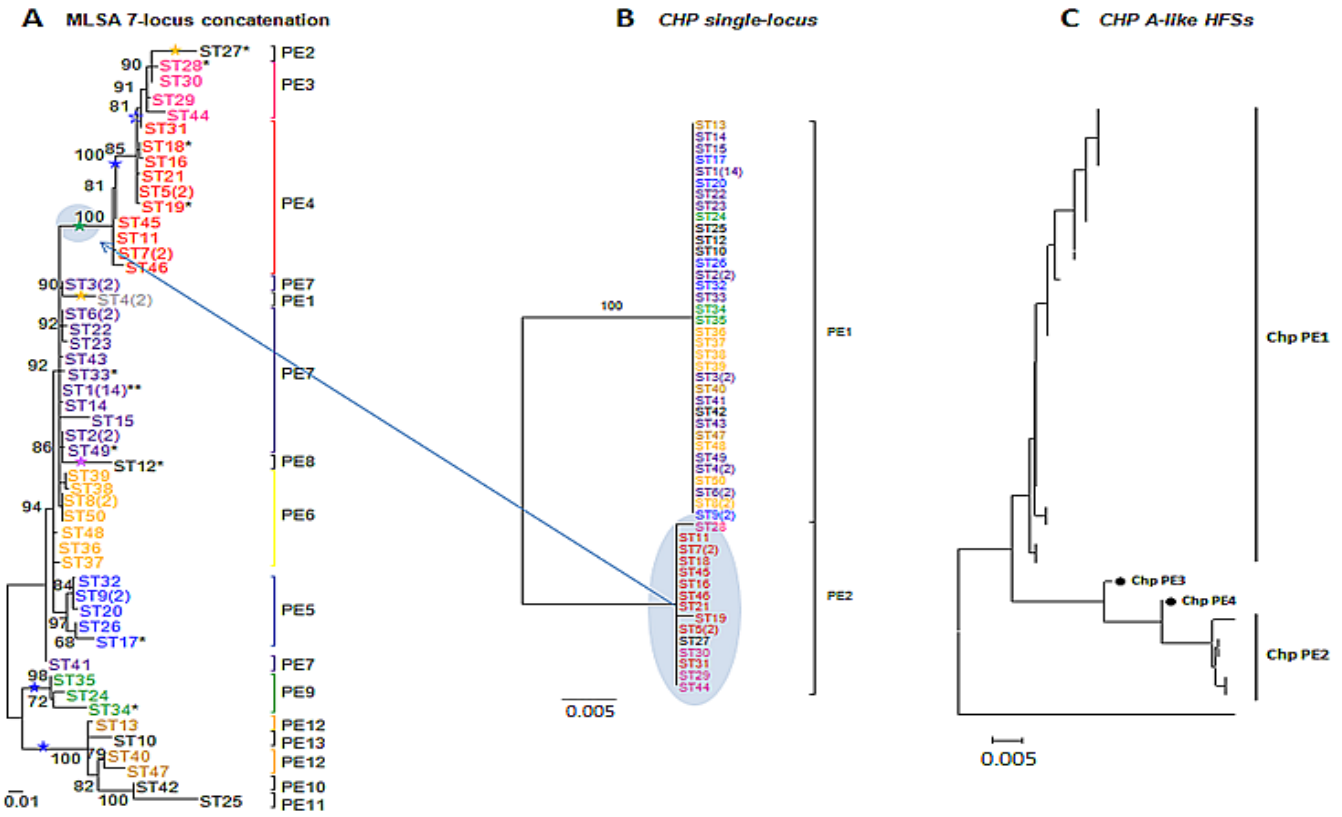


Figure 7.7. Comparison of (A) bacterial artificial chromosome (BAC)-derived multi-locus sequence analysis (MLSA), (B) BAC-derived *chp* single locus analysis (SLA) and (C) barcode-derived high-frequency sequence (HFS) *chp* phylogenies. Arrow matches BAC-derived *chp* SLA putative ecotype (PE) 2 (which corresponds to barcode-derived *chp* HFS PE 2) to the *chp* PE2-specific lineage in the MLSA phylogeny (i.e. all variants that descended from that node (green star) contain *chp* sequences from BAC-derived SLA *chp* PE2). Scale bars correspond to the nucleotide substitutions per site indicated (A and B are reprinted from Melendrez et al., in prep).

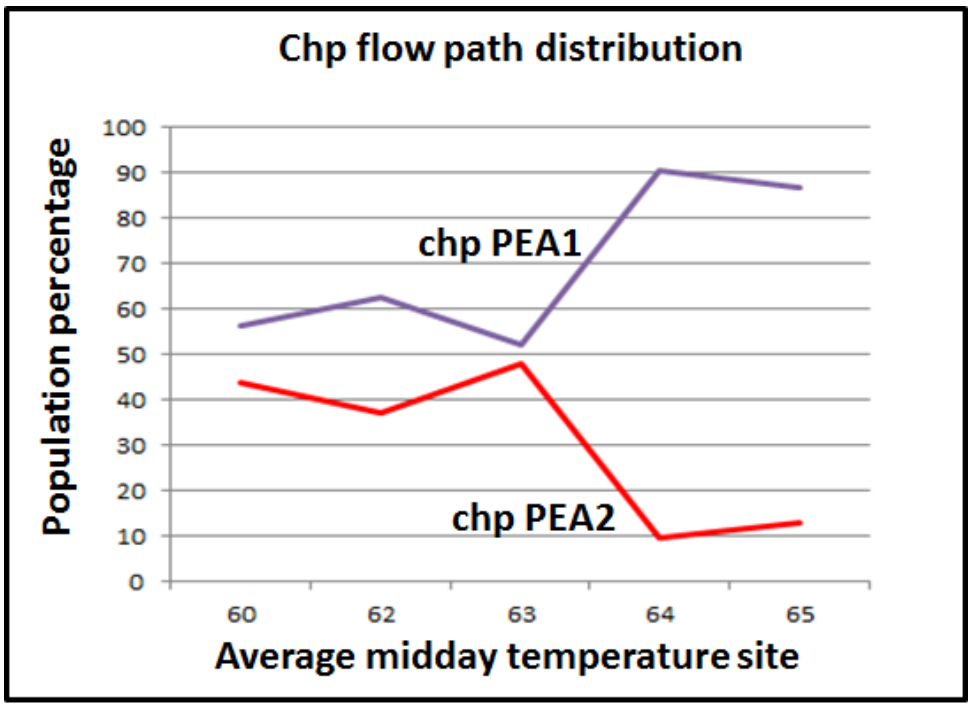


Figure 7.8. Relative abundance of abundant barcode-derived high-frequency sequence *chp*-defined putative ecotypes (PEs) A1 (MLSA PEs 5, 6, 7 and 9) and A2 (MLSA PEs 2, 3 and 4) along the main effluent flow path of Mushroom Spring (MS) at 1°C intervals from 60 to 65°C.

resolving multi-locus analysis are unique ecological species populations. *chp* PE 2 is higher in relative abundance in EZ1 (≤ 63°C), and while *chp* PE 1 is present in EZ1, it is higher in relative abundance in EZ2 (≥64°C). The low molecular resolution of *chp* could not separate individual PEs in SLA, however, the distribution of *chp* PE A2 corresponded to MLSA PEs A3 and A4 *rbsK* SLA distributions, and were higher in relative abundance at lower temperatures (EZ1). The distribution of *chp* PE A1 corresponded to MLSA PE5 distributions, largely increasing in relative abundance above the 63-64°C ecotone boundary, though both were present in relative abundance in EZ1 (60-63°C).

Using *Synechococcus* Genomes to Compare Distributions and Perturbation Responses of PEs Demarcated by MLSA and *psaA*

Rationale

Although the distributions and population dynamics of PEs predicted from *psaA* sequence variation have been studied (Becraft et al, 2011; Becraft et al., submitted; see Chapters 2 and 3), it was impossible to link the *psaA* gene with loci used in the MLSA study because the *psaA* locus is located in a region of the genome not sampled in MLSA analyses. However, the ecotype-specific genomes referred to in the first section of this chapter enabled linkage of these loci. This permitted comparisons of distributions and responses to environmental perturbation of PEs predicted from variation at the *psaA* locus to those of single-loci used in the MLSA analysis.

Materials and Methods

To determine the MLSA PE to which isolates whose genomes that had these sequences corresponded, the 7 loci from the MLSA study of A-like *Synechococcus* BACs were identified by BLAST analyses (Altschul et al., 1990), separated from the genomic databases, aligned, trimmed to match the length of MLSA sequences and concatenated (Melendrez et al., in prep). PEs were demarcated from concatenated MLSA sequences as described in Chapter 3 Materials and Methods. A Neighbor-joining phylogeny containing both MLSA and ecotype-specific genome sequences was constructed using MEGA 5.0 as described in Chapter 4 Materials and Methods. This enabled me to know which *psaA* variants corresponded to which MLSA PE and thereby to compare *psaA* distributions to SLA *rbsK* distributions (Figure 3.3, p.78).

Results and Discussion

Concatenated MLSA loci for genomes representative of *psaA* PEs were grouped by ES analysis together with concatenated BAC-clone MLSA loci, as shown in Figure 7.9. Isolate strains representative of *psaA* PEs A4, A6 and A14 corresponded to MLSA PEs A7 (subclade), A5 and A9, respectively. However, two strains classified as being representative of *psaA* PE A1 were differently distributed in the MLSA phylogeny (Figure 7.9, grey highlighted). This discrepancy with *psaA* PE A1 is highly interesting because it challenges the notion that this PE is a single ecotype. This is consistent with previous observations, wherein PE A1 vertical distributions in EZ2 exhibited multiple peaks of high relative abundance at both the surface and subsurface of mat samples (Chapter 3, Figure 3.3C, p. 78). Additionally, PE A1 appeared to transcribe the *psaA* gene throughout the entire light period in EZ2 (Chapter 3, Appendix Figure B5 B), despite PEs in EZ1 and EZ3 having differential expression patterns of surface and subsurface populations over the diel cycle in relation to light penetration and quality (Chapter 3, Figure 3.4, p. 79 and Appendix B Figure B5).

Distributions of *rbsK* sequence variants of MLSA PEs that corresponded phylogenetically to concatenated genome sequences were analyzed along flow and vertical gradients to examine correspondence to *psaA* defined PEs. Vertical analyses of barcode-derived *rbsK* variants at 63 and 65°C showed that MLSA PE A5 (*psaA* PE A6) and MLSA PE A9 (*psaA* PE A14) are distributed primarily in the deeper mat layers in EZ1, with relative abundances maximizing in the middle mat layers at warmer

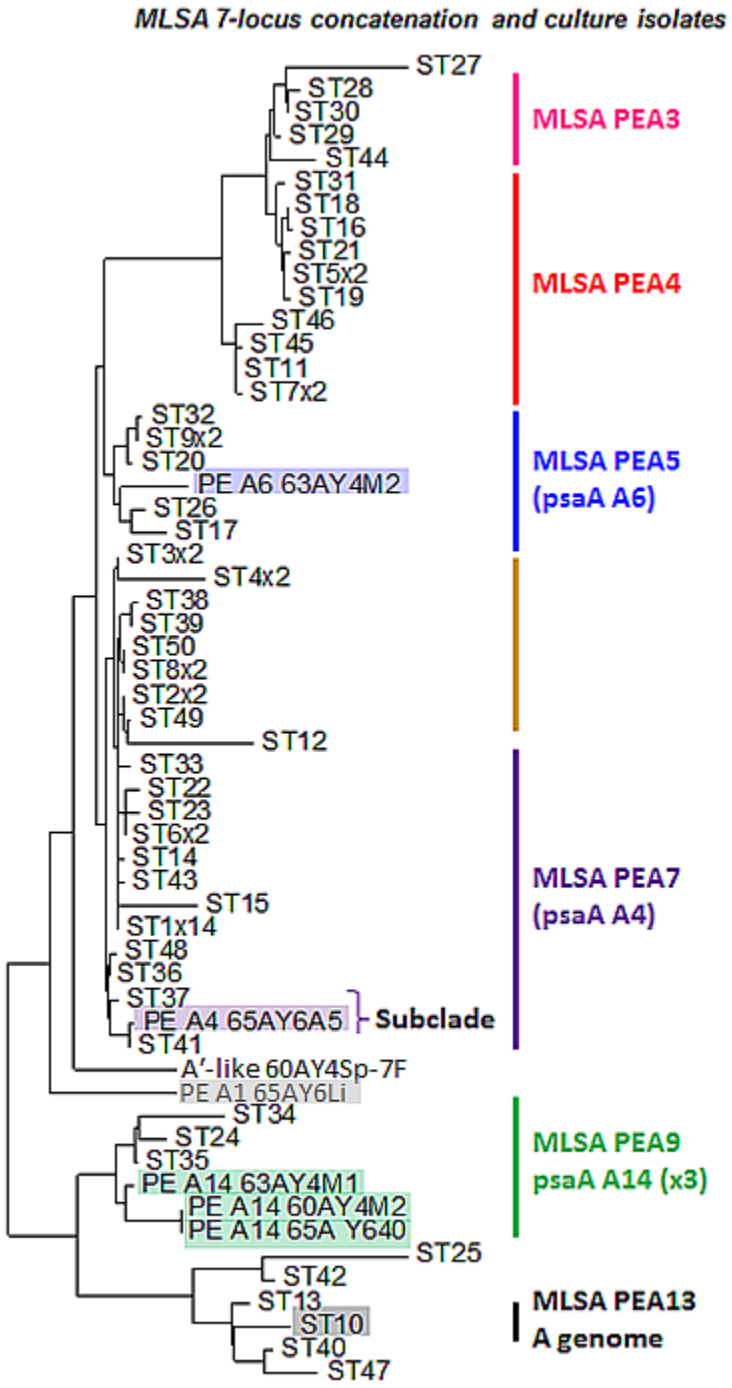


Figure 7.9. 7-locus concatenated MLSA phylogeny based on A-like BAC sequences and sequences from *Synechococcus* isolate strains, showing correspondence between MLSA and *psaA*-defined putative ecotypes (PEs). Isolate strain names are color highlighted to correspond to MLSA PEs.

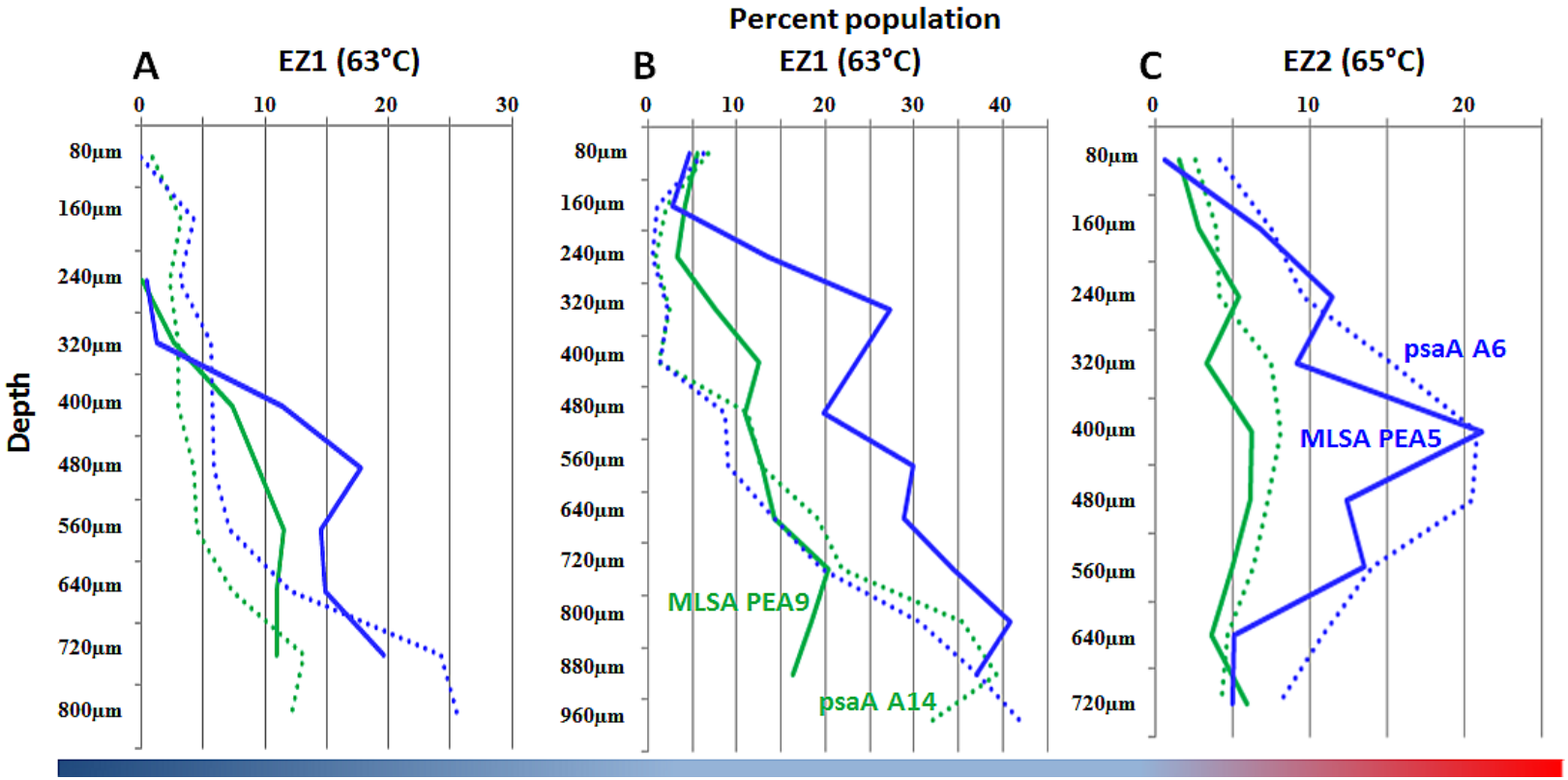


Figure 7.10. Vertical distributions at (A and B) 63°C and (C) 65˚C sites in Mushroom Spring of the relative abundances of *rbsK* variants of multi-locus sequence analysis (MLSA) bacterial artificial chromosome (BAC)-derived PEs A5 and A9 (solid lines) and barcode-derived *psaA* variants representative of PEs A6 and A14 (dotted lines). Missing data points are due to failed sequencing reactions. Colors correspond to the same PE in all parts of the figure.

temperatures (EZ2) (Figure 7.10). Distribution patterns between MLSA PEs (using *rbsK*) and *psaA*-defined PEs that were linked using ecotype-specific genomes corresponded along both temperature and vertical gradients (Chapter 3, Figures 3.2 and 3.3, pp 86 and 87).

In response to ~92% light reduction, MLSA PE A5 declined in relative abundance, and MLSA PE A9 increased in relative abundance at day 6 (Figure 7.11). MLSA PEs A5 and A9 perturbation responses matched those of *psaA* PEs A6 and A14 observed in previous studies, respectively (Chapter 3, Figure 3.5B, p.82).

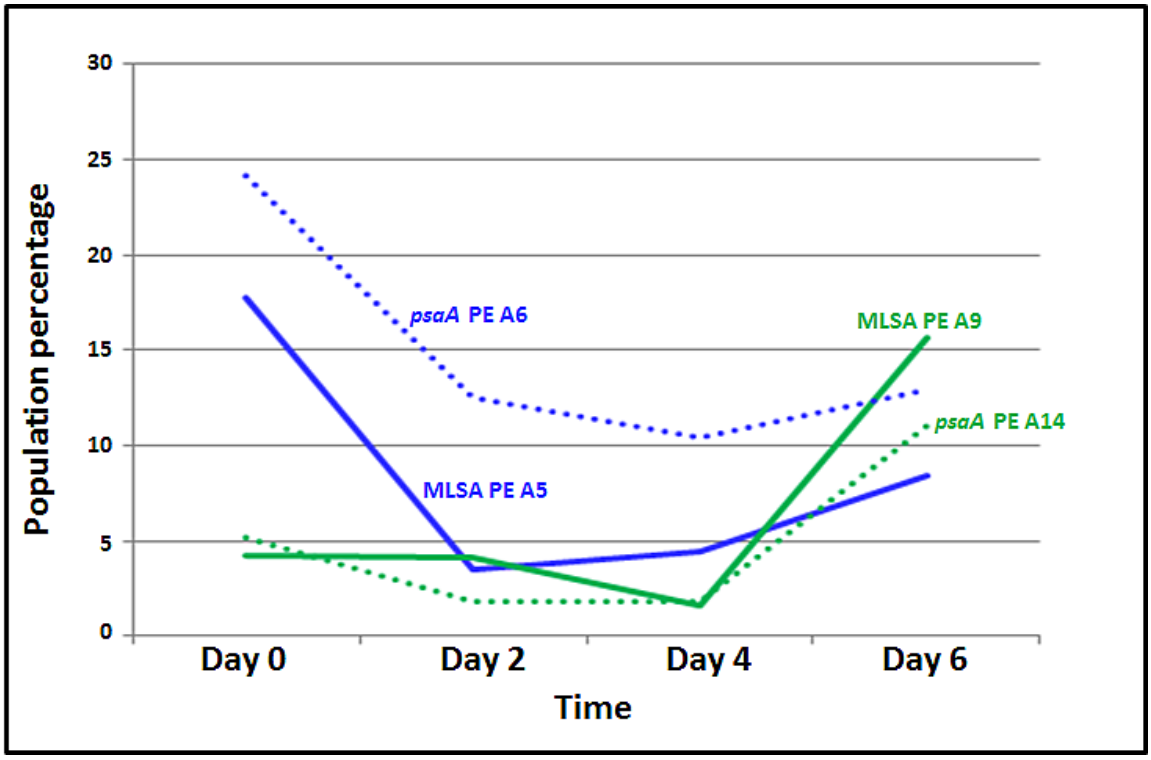


Figure 7.11. Changes in the relative abundances of *rbsK* variants of multi-locus sequence analysis (MLSA) bacterial artificial chromosome (BAC)-derived putative ecotypes (PEs) A5 and A9 and barcode-derived *psaA* variants representative of PEs A6 and A14 in response to ~92% light reduction.

Conclusions

Concatenated genome sequences of *psaA* PEs A6 and A14 corresponded phylogenetically to MLSA PEs A5 and A9, and had similar vertical distributions and responses to light reduction at the *rbsK* locus. However, PE A1 isolates were split phylogenetically into two MLSA lineages, which were too low in relative abundance to accurately track distribution. This could indicate that *psaA* PE A1 consists of more than one ecological population, and that *psaA* lacked the molecular resolution to accurately demarcate ecotypes in this case (also see Chapter 3, Figure 3.3C, p.78 and Appendix B Figure B5).

CHAPTER 8

CONCLUSIONS

Introduction

Microbiologists have struggled to classify microorganisms into species-like units using morphology (Roselló-Mora and Amann, 2001), metabolic properties, DNA hybridization (Wayne et al., 1987) and molecular cutoffs (Konstantinidis and Tiedje, 2007; Konstantinidis, 2011), but often these attempts have proven futile when trying to understand the ecology and evolution of microbial populations in their natural environment (Conner et al., 2010; Hunt et al., 2008; Martiny et al. 2009; Ward, 2006). Multiple studies of diverse microbial habitats have shown the existence of multiple closely related populations that are distributed uniquely along environmental gradients (West and Scanlan, 1999; Béjà et al, 2001; Ward and Cohan, 2005; Cohan et al, 2006; Johnson et al, 2006; Ward et al, 2006; Walk et al, 2007; Hunt et al, 2008; Manning et al, 2008; Martiny et al, 2009; Connor et al, 2010; Denef et al, 2010; Shapiro et al, 2012). These observations have prompted the realization that some microorganisms might exist as ecological species (Ward, 1998; Ward and Cohan, 2005), because they meet one of the basic expectations that ecological species occupy distinct niches. While molecular resolution of the more slowly evolving 16S rRNA locus may be sufficient to discern ancient adaptations, it may be too low to detect younger, more closely related descendant species whose subsequent adaptations require greater molecular resolution to detect (Ferris et al., 2003; Martiny et al, 2009). In an attempt to better understand the ecology

and evolution of natural *Synechococcus* cyanobacterial populations (i.e. ecological species or ecotypes) in Yellowstone National Park, I analyzed a more highly resolving protein-encoding locus across a large number of habitats in space and time. Variation at the *psaA* locus was analyzed using the evolutionary simulation model Ecotype Simulation, which is based on the Stable Ecotype Model of species and speciation, and populations predicted to be ecologically distinct were analyzed to determine if they had the properties expected of ecological species.

Molecular Methods: Past and Present

Initial cloning and sequencing and denaturing gradient gel electrophoresis (DGGE) studies to analyze *Synechococcus* populations using the *psaA* gene showed that putative ecotypes (PEs; populations predicted through evolutionary simulation analysis to be ecologically distinct) embedded within 16S rRNA genotypes (Ferris et al., 1996) had unique flow path and vertical distributions (Chapter 2). However, inherent complications with cloning and sequencing and DGGE prevented me from identifying the distributions of all ecological populations in this environment, or from analyzing other properties that would be expected of ecological species (Cohan, 2011). Cloning and sequencing limits the depth of sampling and the number of habitats that can be analyzed efficiently. Additionally, in DGGE analyses, sequences can have similar melting properties and can co-migrate to locations in the gel so similar that the bands cannot be differentiated as separate genotypes.

Molecular technologies are being developed at a rapid pace, providing microbiologists with the tools to begin to understand the diversity and organization of natural microbial communities. When I began my graduate studies many of these technologies were in their infancy, but during my studies cost-effective, high-throughput 'next-generation' sequencing technologies, in particular Ti454-barcoding and sequencing, became available. One caveat is that Ti454-sequencing generates a large number of errors after strings of identical nucleotides (homopolymers) (Reeder and Knight, 2009), but these errors were corrected by using reference genomes and metagenomic databases, and aligning sequences to the known *psaA* reading frame. Ti454-barcoding and sequencing allowed me to greatly increase sequencing depth, and to apply a sequence-based view of the genetic and ecological diversity within the *Synechococcus* populations simultaneously across multiple habitats.

Properties of Ecological Species

Ti454-barcoding and sequencing allowed for the extension of fine-scale DGGE-based distribution studies, the separation of co-distributed ecological species with similar niches within the local community, the analysis of species-specific patterns of *psaA* gene expression, and the study of unique population dynamics in response to environmental perturbations (Chapter 3). Several abundant PEs declined in relative abundance abruptly and together at certain boundaries, thus demarcating three temperature-defined ecological zones (EZs) along the effluent flow path that were separated by two ecotones (between 63 and 64°C and between 65 and 66°C) (Chapter 3, Figure 3.2, p.77). Vertical

distributions showed that different ecological populations were higher in relative abundance more towards the surface, middle or deeper zones along the vertical aspect of the mat, and positioning was dependent upon the EZ in which a population was located (Chapter 3, Figure 3.3, p.78). Deeper populations were shown to transcribe *psaA* genes at higher light conditions, whereas surface-oriented populations had higher transcript relative abundances during periods of lower irradiance (Chapter 3, Figure 3.4, p.79).

The Stable Ecotype Model of Species and Speciation predicts that ecological species exhibit 1:1 ratios with genetically distinct sequence clusters (at least for genes that are not highly recombinant), are held together by a series of periodic selection events, are made up of ecologically interchangeable individuals, and are ecologically distinct from other populations in the local community (Cohan and Perry, 2007). Variation within ecological populations showed that the majority of ecotypes contained a dominant variant (DV) that made up the plurality of the PE clade, as well as additional high-frequency sequences (HFSs; ≥50 identical representatives across all combined samples) in lesser abundance, and surrounding low-frequency sequence (LFS; <50 copies in all samples) variation. This allowed me to track individual HFSs and surrounding LFS variation within ecological populations. As populations rose or fell in abundance in response to environmental perturbation, the genetic diversity predicted to be contained within an ecotype (e.g. DV, HFSs and LFSs) increased and decreased in proportion with the entire population, indicating that these ecological species are groups of ecologically interchangeable individuals with natural genetic variation (Chapter 3, Figure 3.5, p.82).

Stability of the Local Community.

Contrary to the seasonal stability of 16S rRNA genotypes found by Ferris and Ward (1997), *Synechococcus* ecological populations based on more highly resolving *psaA* variation were shown to change in relative abundance seasonally in mat samples along the effluent flow path, possibly due to the reduction in maximum light intensity and day length in winter months. This could indicate that populations also change seasonally along the vertical aspect of the mat as well, increasing and decreasing in relative abundance at certain depth intervals in response to the changing light environment.

Populations changed in relative abundance over a period of years, possibly in response to disturbance and/or interspecies competition. Additionally, within-population genetic variation (DV and HFSs) also changed to a lesser or greater degree over a period of years, possibly as a result of the forces of periodic selection and/or genetic drifts acting separately upon these populations. Another possibility is that ES is lumping multiple young ecotypes into a single PE, and temporal changes in sequence abundance are actually different ecological populations changing separately from one another. Additionally, these populations could be behaving according to the model of 'speedy speciation' (Cohan and Perry, 2007), where ecological populations are created and go to extinction at rapid rates, which would predictably alter sequence presence and abundance in an environment relatively quickly.

Water flowing above the summertime mat was dominated by A′-like ecotypes that were observed in undisturbed upstream sites, and at lower temperatures, surface-oriented B′-like populations, indicating that dispersal in the water flow is an important factor in

initial mat colonization (i.e. primary succession), and that upstream populations have an initial advantage for recolonizing damaged mat under most environmental conditions (i.e. dispersal is initially more important than adaptation) (Ferris et al., 1997). Colonization of disturbed mat was also influenced by ecological adaptation. Different high-temperature and/or largely surface-oriented populations were able to grow on the recovering mat surface under different light conditions during mat recovery (Chapter 6). Given that mat populations appear to change seasonally in relative abundance, it was not surprising that the cellular content in the water also changed seasonally (Chapter 5), and this in turn should have an impact on population dispersal downstream and colonization of newly forming mat.

Ecological Species Distributions
in the American Northwest.

Expanding on the work of Papke et al. (2003), springs and regions in the American Northwest were observed to have different relative abundances of *Synechococcus* ecological populations, and the relative abundances of PE DVs and HFSs decreased rapidly with geographic distance, suggesting that geographic isolation is important in the evolution of rarely *Synechococcus* populations. Dispersal was apparently limiting enough that populations have rarely migrated among springs across geographic regions, facilitating local adaptation, which appeared to be more important than long-range dispersal in most cases. Ecological populations were shown to have different relative abundances and structures in different springs, and the patterns of genetic

variation predicted to be contained within an ecological species varied among springs within geographic regions.

It cannot be concluded that all clades of sequences predicted by Ecotype Simulation to be ecologically distinct are necessarily representative of unique ecological populations, as they could also be geotypes (i.e. ecologically identical populations that have diverged through geographic isolation; Chapter 4, Figure 4.2, p.95). Ecological (or geographical) populations in the Lower Geyser Basin and West Thumb regions had different within-PE DVs and HFS genetic variation and thus appeared to be evolving separately, despite the relatively small geographic distances separating these basins and the chemical similarity of the springs in these regions.

PEs located at the more distant Mammoth Springs contained unique B′-like genetic variation, which may have been the result of ecological adaptation as well as geographic isolation, since the water chemistry in springs of the Mammoth Terraces group is very different than that of other Yellowstone hot springs (Chapter 4, Figure 4.8, p.111). PEs in Mammoth springs had DV sequences at very high relative abundances compared to HFS and LFS sequence variation in PEs in other springs. Springs in the Mammoth region are ephemeral in nature, and form and dry up over relatively short time spans. This led to the hypothesis that ecotypes at Mammoth Springs have recently experienced bottleneck events (e.g. dispersal to a newly formed spring), and that these ecotypes (geotypes) would contain less associated HFS and LFS variation since they would have had less time after a bottleneck to accumulate variation (Figure 8.1B). This

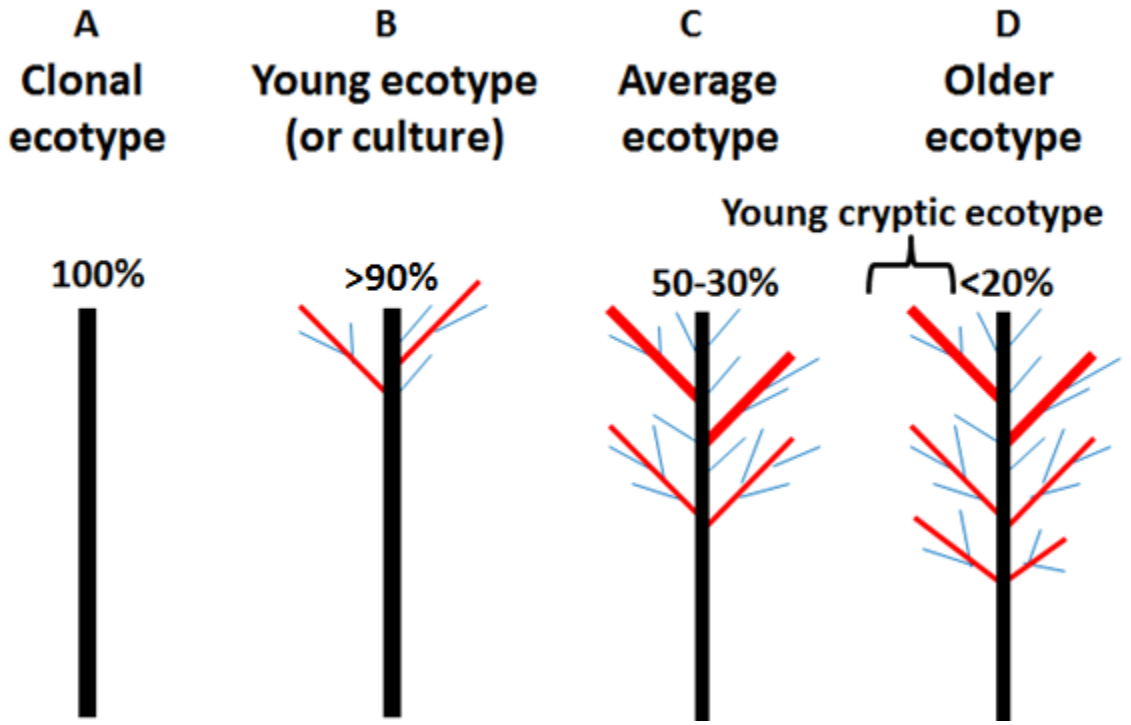


Figure 8.1. Schematic representation of a hypothesized clonal (A), young (B), average (C) and old ecological populations (D). Black line represents the dominant variant (DV) (percent DV is indicated), red lines represent other high-frequency sequences, and blue lines represent surrounding low-frequency sequences. Thickness of lines represents the relative abundance of a sequence variant in the population. Bracket indicates a young, cryptic ecological population contained within a putative ecotype (e.g. PE A1, see below).

approaches the expectation of a clonal unicyanobacterial culture (Table 7.1, p.182 and Figure 8.1A). In contrast, the DVs of older, or more "stable" ecotypes (i.e. more time since dispersal and colonization), would be less relatively abundant as a percentage of the PE population, and the ecotype would contain more relatively abundant associated HFS and LFS variation (see examples below). As mentioned in Chapter 7 (Figure 7.3, p.184), the origin of LFS variation is presently uncertain, but the number and relative abundances

of HFSs is not and should provide an index for the degree to which a PE has diverged, assuming the PE has been correctly demarcated.

Colonization events in Mammoth Springs may have originated with a single genotype, which would be expected if there was minimal dispersal between springs in the Mammoth region, as was observed in the majority of the regions and springs analyzed across the American Northwest. However, despite being limited, dispersal may have been an important factor in determining the founders of a newly-forming spring.

Oregon springs contained unique A-like genetic diversity that was separate from all other springs analyzed in the American Northwest, indicating that dispersal is even more rare over longer distances. However, populations in Oregon shared genetic variation among the chemically similar springs in that region, indicating that dispersal could be important among springs in the same basin in this instance.

This suggests that there is a potential for similar ecotypes in different geographical locations. However, dispersal and recolonization between geographically distant locations appears to be so rare that no cases were observed in which there was >1 phylogenetic cluster per PE.

Additionally, ecological populations were shown to change along environmental gradients in regions and springs in the American Northwest other than those described in Chapter 3. Populations in Octopus Spring and Jack's Spring were shown to change along the effluent flow path where temperature decreases with increasing distance from the source pool, and Clearwater Springs sites with higher or lower pH values were shown to contain different ecological populations in varying relative abundances.

Internal Structures of Ecological Populations.

Ecotype populations were observed to have different structures in different springs and in different years. This could be due to the forces of periodic selection or drift acting upon populations independently, bottleneck events (e.g. the long-range dispersal of a single genotype) or recolonization of disturbed mat. The majority of putative ecotypes observed consisted of varying degrees of within-population genetic variation, and clonal ecological populations were only observed in a few instances (Figure 8.1A; e.g. PE B′9 after UV exclusion and A′-like populations after shifting samples from EZ1 to EZ3; Chapter 3, Figure 3.5C and Appendix D, Figure D2 B, pp. 82 and 291). Mammoth Spring populations may be examples of younger ecotypes, which contained dominant variants that made up the majority of the ecological population and lesser surrounding variation (Figure 8.1B and Appendix F Table F1; e.g. Mam PE B′5 in White Elephant Spring or culture isolate 63AY4M1; Chapter 4, Figure 4.4, p. 99 and Chapter 7, Table 7.1, p. 182). Comparisons between regions and basins in the American Northwest are problematic because multiple basins with different levels of population diversity were combined during analysis. This could erode low level ecological variation around a dominant variant since the most abundant sequences would be preferentially amplified. However, the absence of HFSs >1% in some of the relatively abundant Mammoth Spring populations, and in other populations across the American Northwest (e.g. CW PE B′1; Figures 4.7 and 4.8, pp. 110 and 111), might indicate that these populations are younger. In contrast, most populations in the Lower Geyser Basin and Oregon springs contained multiple HFSs that made up significant proportions of the population (Figure 8.1C; e.g.

PE A12; Figure 3.1, p.75), and a few populations in Mushroom Spring contained a higher than average number of HFSs, possibly indicating a longer lived ecotype (Figure 8.1D; e.g. PE B′8; Figure 3.1, p.75). All sequence variants in populations across the American Northwest contained associated LFS variation. It is unlikely that all of these low abundance sequence variants are from natural random mutations (Drake et al., 1998), and many LFSs associated with HFSs could have been the result of PCR and sequencing error. Further work needs to be done to understand the LFS variation around each HFS, and within-ecotype diversity observations should be constrained to analyzing the number and relative abundance of HFSs demarcated as a putative ecotype.

Ratios of sequence abundance within PEs in individual samples of similar ecology at Mushroom Spring were analyzed to observe changes in ecotype structure over a period of years. Ecotypes in Mushroom Spring between 1996 and 2009 from similar ecologies that made up ≥5% of the community had dominant variant sequences that totaled between 32.8 and 79.5% of the population when analyzed over a period of years (Figure 8.1C and D and Appendix F Table F2). PE B′12 in 2008 and 2009 was possibly the youngest ecotype observed in Mushroom Spring over the period of years analyzed, with the dominant variant sequence making up 70 and 79.5% of the total population in these samples, respectively. It would be unlikely to find a near-clonal population as a result of a periodic selection event given the amount of time between sampling. However, within-population diversity did fluctuate over time, which would be expected if the forces of periodic selection and genetic drift were acting upon ecological populations independently. Additionally, the observed variation in sequence abundance could be the

result of populations forming and going extinct at a rapid pace over time (e.g. speedy speciation model), or predicted ecotypes might contain multiple younger ecological species that *psaA* could not resolve.

Periodic Selection and Ecotype Formation Rates Among Springs

Periodic selection and ecotype formation rates were compared among all springs and studies described in Chapters 2-6 (Table 8.1). With the exception of the 2006 cloning and sequencing study, only HFSs were analyzed due to the enormity of the data sets (between 30,000 and 180,000 sequences) and the constraints of ES at the time datasets were analyzed. In general, ecotype formation events were predicted to be less frequent than periodic selection rates, and both predicted rates were relatively stable within A-like and B′-like *Synechococcus* datasets, as well as within datasets combining A-like and B′-like *Synechococcus* diversity. A striking exception was observed in datasets combining A-like and B′-like based on samples collected over a period of years, which exhibited much higher periodic selection rate estimates. This is interesting, as a greater number of periodic selection events would be expected to have occurred over a longer time period. However, the large confidence intervals (0.2 to 100) observed with periodic selection rate estimates reduce confidence in this inference.

Multiple factors can influence the number of PEs predicted by ES, including the total number of sequences analyzed and number of habitats sampled (Figure 8.2). The number of PEs increased when more sequences were included in the analysis from 18 to

Table 8.1. Summary of periodic selection (PS) and ecotype formation events (EF) rate estimates and the ratio of PS/EF for Mushroom Spring (MS), Octopus Spring (OS), Jack's Spring (JS), Oregon region (Ore), Mammoth region (Mam), Clearwater region (CW) and all B′-like sequences in the biogeographical study of the Northwest United States (NW).

							Biogeography				
	MS A + B′-like 2006	**MS B′-like 2008**	**MS A-like 2008**	**MS A + B′-like 1996-2009**	**MS A + B′-like 1996**	**OS A + B′-like 1989-2005**	**CW A + B′-like 1996**	**JS A-like 1996**	**Ore A-like 1996**	**Mam B′-like 1996**	**NW B′-like 1996**
PS	1.0	1.9	2.7	11.0	0.8	75.0	0.5	0.6	1.7	3.5	0.9
EF	0.2	0.9	0.9	0.6	0.4	0.9	0.1	0.4	1.0	0.9	0.8
PS/EF	5.0	2.0	3.0	18.0	2.0	88.2	5.0	1.5	1.8	3.9	1.1
Ecotypes	18	24	22	29	40	30	21	6	4	5	31
# HFSs	¥	71	47	75	340	118	79	25	21	20	125
# sequences	320	164467	164467	30595	178809	30595	43757	43757	43757	43757	43757
# habitats [a]	3	1	2	2	3	3	2	2	1	1	1

[a] The number of habitats represents the number of ecological zones (EZs) sampled (i.e. temperature defined areas between ecotones).

¥ High-frequency sequences (≥50 identical copies) were not used in the 2006 cloning and sequencing study.

30 with an increase from 320 to 30595 sequences, and from 30 to 40-46 with an increase from 30595 to 164167 sequences. The number of PEs predicted appeared to near an asymptote at ≥164167 sequences, and the further increase in the number of sequences mostly resulted in increased within-PE HFS diversity. The low number of PEs predicted in the 2006 cloning and sequencing study was probably a direct result of fewer total sequences having been analyzed. The increased number of habitats analyzed resulted in more PEs, and PE numbers increased linearly with the increased number of ecologies (e.g., ecological zones) included in the analysis. Results from ES analyses, and the correspondance of data presented in Chapters 3-6, provide support for the inference that most of the abundant PEs in the present day Mushroom Spring environment between ~55-68°C have been identified, unless molecular resolution is still too low to detect very young ecotypes (see below).

Inferring Population Distributions at High Temperature

High-temperature mat samples could not be dissected along the vertical aspect of the mat due to the weak physical structure of the mat samples close to the source pool. However, knowledge of the distributions and expression patterns of ecological species at lower temperatures may provide insight into the distributions of these populations. A′-like *Synechococcus* populations that were capable of recolonization after disturbance (e.g. PEs A′9) were abundant in the water flowing above the mats surface (Chapter 6, Figure 6.6, p.171), which could indicate they are surface populations. However, PE A′8 was also

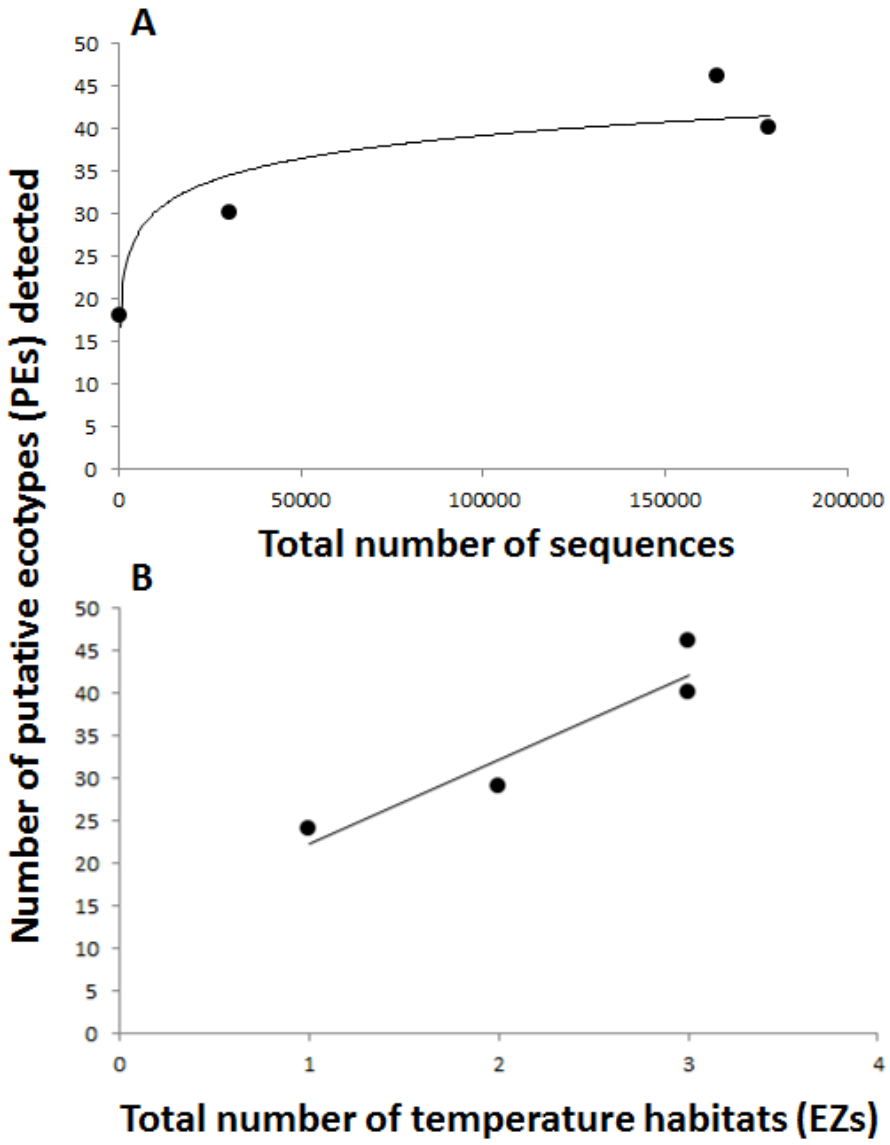


Figure 8.2. The total number of B′-, A- and A′-like putative ecotypes (PEs) as a function of the number of sequences (A) and habitats analyzed (B). Mushroom Spring and Octopus Spring samples were analyzed since they share similar PE diversity and are the most well studied systems. Samples (from left to right) include the 2006 cloning and sequencing study (Chapter 2), the Octopus Spring temporal study (Chapter 5), the 2008 Ti454-barcode study (Chapter 3) and the Mushroom Spring disturbance study (Chapter 6), and all samples contained variation from 3 different temperature-defined habitats (i.e. ecological zones (EZs)). In (B), the 2006 cloning and sequencing samples were excluded due to the low number of sequences analyzed in that study, and the 60°C Mushroom Spring 1996 sample (Chapter 4) was used (far left point). All other points in (B) correspond to part (A).

abundant in the water, but was never seen during initial recolonization of disturbed mat (Chapter 6, Figure 6.3, p.163). Over the diel light transition periods, PE A′9 *psaA* transcript abundance peaked during times of lower irradiance (morning and evening), and PE A′8 peaked in *psaA* transcript abundance closer to midday when light intensities are at their highest levels (Chapter 3, Appendix B Figure B5). This suggests that PE A′9 is surface-oriented, and PE A′8 is possibly located deeper in the mat surface. The looser physical structure of the mat at higher temperatures may facilitate all ecotypes being washed out into the water flowing above the mat surface, and suggests that vertical positioning may have less impact on dispersal at higher temperatures. However, it is possible that PE A′9 had a growth advantage during initial mat recolonization since it is more adapted to environmental conditions at the mat surface.

Temperature and Light as Selective Forces

Ecotypes rarely crossed upper temperature ecotone boundaries from a lower EZ to a higher EZ (e.g. B′-like diversity crossing into EZ2, or A-like diversity crossing into EZ3), and PE overlap was more frequent at lower temperatures (e.g. more A-like diversity in EZ1 than A′-like diversity in EZ2). This could be because adaptation to higher temperatures is more difficult, and the upper temperature ecotone boundary is more difficult to cross than the lower-temperature ecotone boundary. This could also blur the transition between EZ1, and, another lower-temperature EZ. More recently evolved ecotypes within ecotone boundaries distribute themselves along the vertical aspect of the mat, where light quantity and quality vary with depth and dynamic chemical gradients

differ. Vertical distributions also appear to be dependent upon the EZ in which the PE is located. The same *psaA*-based PEs are closer to the mat surface at warmer temperatures, which could be explained by ecotypes requiring higher light intensities to reach the compensation point for photosynthesis at higher temperatures. Alternatively, if *psaA* offers too little molecular resolution to detect young species, the population distribution could be a consequence of the combination of different species inhabiting different vertical intervals at different temperatures and interspecies competition affecting populations differently in separate EZs. Interestingly, while the problem of molecular resolution is similar, in the case of marine *Prochlorococcus* light adaptation appears to be more ancient than temperature adaptation (Martiny et al. 2009).

Other Ti454-barcoding Applications

The expected clonality of laboratory strains allowed me to use Ti454-barcoding to discern whether *Synechococcus* isolates representative of native PEs were unicyanobacterial, or if they contained multiple populations from the environment. Since this technique was implemented, with the collaborative efforts of Millie Olsen and Shane Nowack, we have obtained genome sequences of 8 isolates (in addition to those of the A (JA-3-3Aa) and B′ (JA-2-3B′a (2-13)) strains, which had been obtained previously) that are representative of different PEs, and we have shown that sequenced isolate strains representative of some ecological populations contain ecotype-specific genes and that cultures representative of ecological populations have light adaptations that correlate with their *in situ* vertical distributions (Nowack et al., in prep).

Barcode analysis of the distributions of *rbsK* and *chp* sequences used in multi-locus sequence analyses (MLSA) demonstrated that insufficient molecular resolution and/or recombination can complicate identification of ecological species populations (Melendrez et al., in prep). Additionally, in cases where concatenated genome sequences corresponded phylogenetically with MLSA-defined PEs, I was able to compare the distributions of *rbsK* and *chp* variants representative of MLSA-defined PEs to distributions of *psaA*-defined PEs. The distribution of *psaA*-defined PEs A6 and A14 were shown to match to those of the corresponding MLSA-based PEs, and the similarity of these distributions and responses to environmental perturbation provided further confidence that these PEs predicted using the *psaA* locus are representative of ecologically distinct populations (Chapter 7, Figures 7.10 and 7.11, pp. 201 and 202).

Unresolved Ecological Populations

There were cases in which putative ecotypes could not be shown to be ecologically distinct. For example, in Chapter 3 PEs A4 and A6 could not be shown to be ecologically distinct based on distribution patterns and responses to environmental perturbations alone. However, seasonal distributions provided preliminary evidence that PE A4 is an ecologically distinct, low-light adapted population that is more relatively abundant in winter, whereas PE A6 is more relatively abundant in summer (Chapter 5, Figure 5.3, p. 135).

The case of PE A1 is particularly interesting, as multiple lines of evidence suggest that this PE groups multiple young ecotypes (Chapter 2, Table 2.2, p.56). First, in

Chapter 3 PE A1 was shown to be relatively abundant throughout the entire upper green mat layer in EZ2 (Figure 3.3C, p.78), but the pattern along the vertical aspect shows a bi-modal distribution, and with peaks in relative abundance near the surface and deeper in the mat green layer. Second, the continuous *psaA* transcription pattern throughout the day in EZ2 (Appendix B Figure B5) contrasts with the mid-day patterns observed for populations residing deeper in the mat at EZ1 (and possibly at EZ3). This could be because PE A1 is actually two populations with different depth distributions that are lumped into one PE whose transcripts cannot be distinguished in the *psaA* sequence region studied. Third, MLSA demonstrated that the genomes of two *psaA*-based PE A1 strains are phylogenetically distinct in MLSA analyses (Chapter 7; Figure 7.9, p.200). The *psaA* locus might not differentiate between PEs embedded within PE A1, possibly because these hypothesized PE A1-like ecological populations are younger, and *psaA* lacks the required molecular resolution to accurately demarcate them. An example of this was observed in the case of *rbsK* analyses, where MLSA analyses split apart sequences that had been lumped by single-locus analysis (Chapter 7, Figure 7.4, p.190). As a result, it is possible that closely related A-like PEs that are present at multiple temperatures and depth intervals are actually representative of different ecological populations unresolved by using *psaA* sequences to demarcate PE A1. Alternatively, PE A1 could be a generalist population, capable of inhabiting multiple EZs at different depths, and is able to acclimate and transcribe the *psaA* gene at different times of day at different temperatures. Additionally, it could mean that the Stable Ecotype Model hypothesis must be rejected in

the case of PE A1, possibly in favor of the speedy speciation model (Cohan and Perry, 2007).

A Warning About the *Synechococcus* sp. Strain B′ Genome

In analyzing the genetic diversity of isolate strains, the B′ genome (JA-2-3B′a (2-13)) *psaA* sequences were found to correspond to 3 DVs of phylogenetically distant PEs, two of which were identical to the DVs of different PEs detected in the environment (Table 7.1), and the third of which was not present in Mushroom Spring. These results indicate that the B′ genome described in Bhaya et al. (2007) is possibly not representative of an individual within an ecological species, but is actually an amalgam of the genomes of different individuals representative of multiple ecological species in the environment (also see Chapter 7, Figure 7.2, p.180).

A Warning About Ti454 Multiplex sequencing

Failed sequencing reactions resulted from attempting to simultaneously sequence multiple loci with different primers on the same sequencing plate (multiplex sequencing). Differences in the energetics of binding of different primers may have biased annealing resulting in biased sequencing (Chapter 7, Appendix E Table E1). Sequencing failure was also observed in a collaborative NASA project to analyze the distribution of other mat phototrophs. In that study, primers for different target genes in different phototrophic populations were used in multiplex sequencing, and as with the MLSA-locus distribution

study, failed sequencing reactions were more common than successful ones. It is recommended that future barcoding research focus on a specific locus (and one primer) for each individual sequencing run.

Concluding Remarks and Future Directions

The application of Ti454-barcoding and sequencing enabled me to expand upon previous research, validating the distinctness of ecological populations predicted by ES. Increased habitat variation and sequencing depth allowed me to confirm that many of the abundant *Synechococcus* PEs predicted from evolutionary simulation analyses of *psaA* sequence variation were ecologically distinct populations, containing ecological homogeneous individuals. The deep sequencing provided by Ti454 technology also allowed me to better understand ecological species distributions across space and time.

Overall, *Synechococcus* ecological species appear to be distributed differently on local and regional spatial scales, to change on short- and long-term temporal scales and in response to disturbance, and to behave according to the evolutionary and ecological rules that govern plant and animal species.

Now that we have a better understanding of *Synechococcus* (i) fine-scale distributions of ecological populations, (ii) species-specific expression patterns, (iii) population-specific responses to environmental perturbations, (iv) large-scale biogeographic ecological species distributions, (v) temporal stability within local communities, and (vi) initial colonization in response to environmental disturbance, the future of the Ward Lab includes many interesting possibilities. The acquisition of

ecotype-specific genomes within and between multiple ecologically distinct populations, and the ability to test the light and temperature (and other) adaptations of ecotypes through growth and competition experiments, will greatly advance our understanding of *Synechococcus* populations within these local communities. I am humbled to have helped advance the understanding of the ecology, evolution and population structure of thermophilic *Synechococcus* cyanobacteria in Yellowstone National Park during my time at Montana State University while in the Ward Lab.

REFERENCES CITED

Allewalt, J. P., M. M. Bateson, N. P. Revsbech, K. Slack and D. M. Ward (2006). "Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in Octopus Spring microbial mat community of Yellowstone National Park." *Applied and Environmental Microbiology* **72**:544-550.

Altschul S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." *Journal of Molecular Biology* **215**:403-410.

Amann, R. I., W. Ludwig and K. H. Schleifer (1995). "Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation." *Microbiology Review* **59**:143-169.

Andersson, A. F., M. Lindberg, H. Jakobsson, F. Bäckhed, P. Nyrén and L. Engstrand (2008). "Comparative analysis of human gut microbiota by barcoded pyrosequencing." *PLoS ONE* **3**:e2836.

Bass-Becking, L. (1934). Geobiologie of inleiding tot de milieukunde. Van Stockum & Zoon, The Hague, the Netherlands.

Bauld, J. and T. D. Brock (1973). "Ecological studies of *Chloroflexis*, a gliding photosynthetic bacterium." *Archives of Microbiology* **92**:267-284.

Becraft, E. D., F. M. Cohan, M. Kühl, S. I. Jensen and D. M. Ward (2011). "Fine-scale distribution patterns of *Synechococcus* ecological diversity in microbial mats of Mushroom Spring, Yellowstone National Park." *Applied and Environmental Microbiology* **77**:7689-7697.

Becraft, E. D., F. M. Cohan, M. Kühl, S. I. Jensen, D. B. Rusch, J. M. Wood and D. M. Ward. "Pyrosequencing analyses of *Synechococcus* ecological species inhabiting the microbial mat of Mushroom Spring, Yellowstone National Park." Submitted.

Béjà, O., E. N. Spudich, J. L. Spudicn, M. Leclerc & E. F. DeLong (2001). "Proteorhodopsin phototrophy in the ocean." *Nature* **411**:786-789.

Bhaya, D., A. R. Grossman, A. S. Steunou, N. Khuri, F. M. Cohan, N. Hamamura, M. C. Melendrez, M. M. Bateson, D. M. Ward and J. F. Heidelberg (2007). "Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analysis." *ISME Journal* **1**:703-713.

Boucher, Y., C. J. Douady, R. T. Papke, D. A. Walsh, M. E. R. Boudreau, C. L. Nesbø, R. J. Case and W. F. Doolittle (2003). "Lateral gene transfer and the origins of prokaryotic groups." *Annual Review Genetics* **37**: 283-328.

Brock, T. D. (1978). Thermophilic microorganisms and life at high temperatures. Springer Verlag, New York.

Brock, T. D. and M. L. Brock (1968). "Relationship between environmental temperature and optimum temperature of bacteria along a hot spring thermal gradient." *Journal of Applied Bacteriology* **31**:54-58.

Cohan, F. M. (2002). "What are bacterial species?" *Annual Review Microbiology* **54**: 457-487.

Cohan, F. M. (2006). "Towards a conceptual and operational union of bacterial systematics, ecology and evolution." *Philosophical Transactions of the Royal Society B* **361**:1985-1996.

Cohan, F. M. (2011). "Are species cohesive? A view from bacteriology." *In* Bacterial Population Genetics: A Tribute to Thomas S. Whittam. Walk S. and P. Feng (eds.). American Society for Microbiology Press, Washington, pp. 43-65.

Cohan, F. M. and A. F. Koeppel (2008). "The origins of ecological diversity in prokaryotes." *Current Biology* **18**:1024-1034.

Cohan, F. M., A. Koeppel and D. Krizanc (2006). "Sequence-based discovery of ecological diversity within *Legionella*." *In Legionella*: state of the art 30 years after its recognition. N. P. Cianciotto, Y. A Kwaik, P. H. Edelstein, B. S. Fields, D. F. Geary, T. G. Harrison, C. A. Joseph, R. M. Ratcliff, J. E. Stout, and M. S. Swanson (eds). ASM Press, Washington, DC, pp. 367-376.

Cohan, F. M. and E. B. Perry (2007). "A systematics for discovering the fundamental units of bacterial diversity." *Current Biology* **17**:373-386.

Connor, N., J. Sikorski, A. P. Rooney, S. Kopac, A. F. Koeppel, A. Burger, S. G. Cole, E. B. Perry, D. Krizanc, N. C. Field, M. Slaton and F. M. Cohan (2010). "Ecology of speciation in the genus *Bacillus*." *Applied and Environmental Microbiology* **76**: 1349-1358.

Coppola, S., G. Mauriello, M. Aponte, G. Moschetti and F. Villani (2000). "Microbial succession during ripening of Naples-type salami, a southern Italian fermented sausage." *Meat Science* **56**:321-329.

Darwin, C. R. (1859). The Origin of Species. Harvard Classics, Vol 11. John Murray, London.

Denef, V. J., L. H. Kalnejais, R. S. Mueller, P. Wilmes, B. J. Baker, B. C. Thomas, N. C. VerBerkmoes, R. L. Hettich and J. F. Banfield (2010). "Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities." *Proceedings of the National Academy of Science* **107**:2383-2390.

de Bruyn, J., F. C. Boogerd, P. Bos and J. G. Kuenen (1990). "Floating filters, a novel technique for isolation and enumeration of fastidious, acidophilic, iron-oxidizing, autotrophic bacteria." *Applied and Environmental Microbiology* **56:**2891-2894.

de Queiroz, K. (2005). "Ernst Mayr and the modern concept of species." *Proceedings of the National Academy of Science* **102**:6600-6607.

de Queiroz, K. (2007). "Species concepts and species delimitation." *Systems Biology* **56**: 879-886.

Doolittle, W. F. and R. T. Papke (2006). "Genomics and the bacterial speciation problem." *Genome Biology* **7**:116.1-116.7.

Drake, J. W., B., Charlesworth, D., Charlesworth and J. F. Crow (1998). "Rates of spontaneous mutation." *Genetics* **148**:1667-1686.

DuRand, M. D., R. J. Olson and S. W. Chisholm (2001). "Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea." *Deep Sea Research Part II: Topical Studies in Oceanography* **48**:1983-2003.

Dykhuizen, D. E. (1998). "Santa Rosalia revisited: Why are there so many species of bacteria?" *Antonie van Leeuwenhoek* **73**:25-33.

Falush, D., M. Stephens and J. K. Pritchard (2003). "Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies." *Genetics* **164**:1567-1587.

Fenchel, T. and B. J. Finlay (2004). "The Ubiquity of Small Species: Patterns of Local and Global Diversity." *BioScience* **54**:784-777.

Ferris, M. J., M. Kühl, A. Wieland and D. M. Ward (2003a). "Different light-adapted ecotypes in a 68ºC *Synechococcus* mat community revealed by analysis of 16S-23S intervening transcribed spacer variation." *Applied and Environmental Microbiology* **69**:2893-2898.

Ferris, M. J., M. Kühl, A. Wieland and D. M. Ward (2003b). Cyanobacterial ecotypes in different optical microenvironments of a 68^{o}C hot spring mat community revealed by 16S-23S rRNA internal transcribed spacer region variation. *Applied and Environmental Microbiology* **69**:2893-2898.

Ferris, M. J., G. Muyzer and D. M. Ward (1996a). "Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community." *Applied and Environmental Microbiology* **62**:340-346.

Ferris, M. J., S. C. Nold, N. P. Revsbech and D. M. Ward (1997). "Population structure and physiological changes within a hot spring microbial mat community following disturbance." *Applied and Environmental Microbiology* **63**:1367-1374.

Ferris, M. J., A. L. Ruff-Roberts, E. D. Kopczynski, M. M. Bateson and D. M. Ward (1996b). "Enrichment culture and microscopy conceal diverse thermophilic *Synechococcus* populations in a single hot spring microbial mat community." *Applied and Environmental Microbiology* **62**:1045-1050.

Ferris, M. J. and D. M. Ward (1997). "Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis." *Applied and Environmental Microbiology* **63**:1375-1381.

Fierer, N. (2008). "Microbial biogeography: patterns in microbial diversity across space and time." *In* Accessing uncultivated microorganisms: from the environment to organisms and genomes and back. K. Zengler (ed). ASM Press, Washington DC, pp. 95-115.

Francisco, J. C., F. M. Cohan and D. Krizanc (2012). "Demarcation of bacterial ecotypes from DNA sequence data: a comparative analysis of four algorithms." 2nd IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), Las Vegas.

Fries, M. R., G. D. Hopkins, P. L. McCarty, L. J. Forney and J. M. Tiedje (1997). "Microbial succession during a field evaluation of phenol and toluene as the primary substrates for trichloroethene cometabolism." *Applied and Environmental Microbiology* **63**:1515-1522.

Gevers, D., F. M. Cohan, J. G. Lawrence, B. G. Spratt, T. Coenye, E. J. Feil, E. Stackenbrandt, Y. Van de Peer, P. Vandamme, F. L. Thompson and J. Swings (2005). "Opinion: Re-evaluating prokaryotic species." *Nature Reviews Microbiology* **3**:733-739.

Giovannoni, S. J., T. B. Britschgi, C. L. Moyer and K. G. Field (1990). "Genetic diversity in Sargasso Sea bacterioplankton." *Nature* **345**:60-63.

Giovannoni, S. J. and K. L. Vergin (2012). "Seasonality in ocean microbial communities." *Science* **335**:671-676.

Goodfellow, M., G. P. Manfio and J. Chun (1997). "Towards a practical species concept for cultivable bacteria." *In* Species: the units of biodiversity. Claridge M. F., H. A. Dawah and M. R. Wilson (eds.). Chapman and Hall, London, pp. 25-59.

Head, I. M., J. R. Saunders and R. W. Pickup (1998). "Microbial evolution, diversity and ecology: a decade of ribosomal RNA analysis of uncultivated microorganisms." *Microbial Ecology* **35**:1-21.

Hugenholtz, P., B. M. Goebel and N. R. Pace (1998). "Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity." *Journal of Bacteriology* **180**:4765-4774.

Hughes-Martiny, J. B., J. M. Brendan, J. H. Bohannan, R. K. Brown, J. A. Colwell, J. L. Fuhrman, M. Green, C. Horner-Devine, M. Kane, J. Adams-Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Øvreås, A. L. Reysenbach, V. H. Smith and J. T. Staley (2006). "Microbial biogeography: putting microorganisms on the map." *Nature Reviews* **4**:102-122.

Hunt, D. E., L. A. David, D. Gevers, S. P. Preheim, E. J. Alm and M. F. Polz (2008). "Resource partitioning and sympatric differentiation among closely related bacterioplankton." *Science* **320**:1081-1085.

Ishii, K., M. Fukui and S. Takii (2000). "Microbial succession during a composting process as evaluated by denaturing gradient gel electrophoresis analysis." *Journal of Applied Microbiology* **89**:768-777.

Jensen, S. I., A-S. Steunou, D. Bhaya, M. Kühl and A. R. Grossman (2011). "*In situ* dynamics of O_2, pH and cyanobacterial transcripts associated with CCM, photosynthesis and detoxification of ROS." *ISME Journal* **5**: 317-328.

Johnson, Z. I., E. R. Zinser, A. Coe, N. P. McNulty, E. Malcolm, S. Woodward and S. W. Chisholm (2006). "Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients." *Science* **311**:1737-1740.

Kaštovská, K., J. Elster, M. Stibal and H. Šantrůčková (2005). "Microbial assemblages in soil microbial succession after glacial retreat in Svalbard (high arctic)." *Microbial Ecology* **50**:396-407.

Klatt, C. G., J. W. Wood, D. B. Rusch, M. M. Bateson, N. Hamamura, J. F. Heidelberg, A. R. Grossman, D. Bhaya, F. M. Cohan, M. Kühl, D. A. Bryant and D. M. Ward (2011). "Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential." *ISME Journal* **5**:1262-1278.

Koch, A. L. (1973). "The pertinence of the periodic selection phenomenon to prokaryote evolution." *Genetics* **77**:127-142.

Koeppel, A., E. B. Perry, J. Sikorski, D. Krizanc, A. Warner, D. M. Ward, A. P. Rooney, E. Brambilla, N. Connor, R. M. Ratcliff, E. Nevo and F. M. Cohan (2008). "Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics." *Proceedings of the National Academy of Sciences* **105**:2504-2509.

Koeppel, A. F., J. O. Wertheim, L. Barone, N. Gentile, D. Krizanc and F. M. Cohan (2013). "Speedy speciation in a bacterial microcosm: new species can arise as frequently as adaptations within a species." *ISME Journal* doi: 10.1038/ ismej. 2013.3.

Konstantinidis, K. T. (2011). "Metagenomic insights into bacterial species." *In* Handbook of molecular and microbial ecology, volume 1: metagenomics and complementary approaches. Frans J. de Bruijn (ed.), pp. 89-98.

Konstantinidis, K. T., and J. M. Tiedje (2005). "Genomic insights that advance the species definition for prokaryotes." *Proceedings of the National Academy of Sciences U.S.A.* **102**: 2567-2572.

Konstantinidis, K. T. and J. M. Tiedje (2007). "Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead." *Current Opinions in Microbiology* **10**:504-509.

Kühl, M. (2005) "Optical microsensors for analysis of microbial communities." *Methods in Enzymology* **397**: 166-199.

Larkin, M. A., G. Blackshield, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." *Bioinformatics* **23**:2947-2948.

Lincoln, R., G. Boxshall and P. Clark (1998). A dictionary of ecology, evolution and systematics (2nd edition). Cambridge University Press (eds). Cambridge, United Kingdom.

Liu, Z., C. G. Klatt, M. Ludwig, D. B. Rusch, S. I. Jensen, M. Kühl, D. M. Ward and D. A. Bryant (2012). "'*Candidatus* Thermochlorobacter aerophilum': an aerobic chlorophotoheterotrophic member of the phylum Chlorobi defined by metagenomics and metatranscriptomics." *ISME Journal* **6**:1869-1882.

Liu Z., C. G. Klatt, J. M. Wood, D. B. Rusch, M. Ludwig, N. Wittekindt, L. P. Tomsho, S. C. Schuster, D. M. Ward and D. A. Bryant (2011). "Metatranscriptomic analyses of chlorophototrophs of a hot spring microbial mat." *ISME Journal* **5**:1279-1290.

Manning, S. D., A. S. Motiwala, A. C. Springman, W. Qi, D. W. Lacher, L. M. Ouellette, J. M. Mladonicky, P. Somsel, J. T. Rudrik, S. E. Dietrich, W. Zhang, B. Swaminathan, D. Alland and T. S. Whittam (2008). "Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks." *Proceedings of the National Academy of Sciences U.S.A.* **105**:4868-4873.

Martin D. P., P. Lemey, M. Lott, V. Moulton, D. Posada and P. Lefeuvre (2010). "RDP3: a flexible and fast computer program for analyzing recombination." *Bioinformatics* **26**:2462-2463.

Martiny, A. C., A. P. K. Tai, D. Veneziano, F. Primeau and S. W. Chisholm (2009). "Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*." *Environmental Microbiology* **11:**823-832.

Mayden, R. L. (1997). "A hierarchy of species concepts: the denouement in the saga of the species problem." *In* Species: the units of biodiversity. Claridge M. F., H. A. Dawah and M. R. Wilson (eds). Chapman and Hall: London, pp. 381-424.

Mayr, E. (1942). Systematics and the origin of species. Columbia University Press: New York.

Mayr, E. (1982). The Growth of Biological Thought: Diversity, Evolution, and Inheritance. Belknap Press of Harvard University Press, Cambridge, MA.

McDougald, D., S. A. Rice, N. Barraud, P. D. Steinberg and S. Kjelleberg (2012). "Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal." *Nature Reviews Microbiology* **10**:39-50.

Meeks, J. C. and R. W. Castenholz (1971). "Growth and photosynthesis in an extreme thermophile, *Synechococcus lividus* (Cyanophyta)." *Archives of Mikrobiology* **78**:25-41.

Melendrez, M. C., J.M. Wood, E. D. Becraft, F. M. Cohan, J.F. Heidelberg, D. Rusch and D. M. Ward. "The relative importance of recombination and selection on the evolution of *Synechococcus* species inhabiting a hot spring cyanobacterial mat." (in preparation).

Melendrez, M. C., R. K. Lange, F. M. Cohan and D. M. Ward (2011). "Influence of molecular resolution on sequence-based discovery of ecological diversity among *Synechococcus* populations in an alkaline siliceous hot spring microbial mat." *Applied and Environmental Microbiology* **77**:1359-1367.

Mehrdad H., G. A. C. Singer, E. L. Clare and P. D. N. Hebert (2007). "Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring." *BMC Biology* **5:**24.

Miller, S. R. and R. W. Castenholz (2000). "The evolution of thermotolerance in hot spring cyanobacteria of the genus *Synechococcus*." *Applied and Environmental Microbiology* **66**:4222-4229.

Miller, S. R., M. D. Purugganan and S. E. Curtis (2006). "Molecular population genetics and phenotypic diversification of two populations of the thermophilic *Masigocladus laminosus*." *Applied and Environmental Microbiology* **72**:2793-2800.

Miller, S. R., C. E. Winguard and R. W. Castenholz (1998). "Effects of visible light and UV radiation on photosynthesis in a population of a hot spring cyanobacterium, a *Synechococcus* sp., subjected to high-temperature stress." *Applied and Environmental Microbiology* **64**:3893-3899.

Morin, P. J. (2011). Community ecology (2nd edition). Wiley. Hoboken, NJ.

Mueller, R. S., B. D. Dill, C. Pan, C. P. Belnap, B. C. Thomas, N. C. VerBerkmoes and J. F. Banfield (2011). "Proteome changes in the initial bacterial colonist during ecological succession in an acid mine drainage biofilm community." *Environmental Microbiology* **13**:2279-2292.

Muramatsu, M., K. Sonoike and Y. Hihara (2009). "Mechanism of downregulation of photosystem I content under high-light conditions in the cyanobacterium *Synechocystis* sp. PCC 6803." *Microbiology* **155**:989-996.

Muyzer, G., E. C. Dewaal and A. G. Uitterlinden (1993). "Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S ribosomal RNA." *Applied and Environmental Microbiology* **59**:695-700.

Noguez, A. M., H. T. Arita, A. E. Escalante, L. J. Forney, F. García-Oliva and V. Souza (2005). "Microbial macroecology: highly structured prokaryotic soil assemblages in a tropical deciduous forest." *Global Ecology and Biogeography* **14**:241-248.

Norris, T. B., T. R. McDermott and R. W. Castenholz (2002). "The long-term effects of UV exclusion on the microbial composition and photosynthetic competence of bacteria in hot-spring microbial mats." *FEMS Microbiology Ecology* **39**:193-209.

Nowack, S., M. Olsen, E. D. Becraft, D. Bryant and D. M. Ward. "Phenotypic response with respect to temperature and light of *psaA* isolates of *Synechococcus* spp. from Mushroom Spring, Yellowstone National Park." (in preparation).

Nübel, U., M. M. Bateson, V. Vandieken, A. Wieland, M. Kühl and D. M. Ward (2002). "Microscopic examination of distribution and phenotypic properties of phylogenetically diverse *Chloroflexaceae*-related bacteria in hot spring microbial mats." *Applied and Environmental Microbiology* **68**:4593-603.

Papke, T. R., N. B. Ramsing, M. M. Bateson and D. M. Ward (2003). "Geographical isolation in hot spring cyanobacteria." *Environmental Microbiology* **5**:650-659.

Papke, R. T., O. Zhaxybayeva, E. Feil, K. Sommerfeld, D. Muise and W. F. Doolittle (2007). "Searching for species in haloarchaea." *Proceedings of the National Academy of Science* **104**:14092-14097.

Quince C., A. Lanzen, T. P. Curtis, R. J. Davenport, N. Hall. I. M. Head, L. F. Read and W. T. Sloan (2009). "Accurate determination of microbial diversity from 454 pyrosequencing data." *Nature Methods* **6**:639-641.

Quince C., A. Lanzen, R. J. Davenport, P. J. Turnbaugh (2011). "Removing noise from pyrosequenced amplicons." *BMC Bioinformatics* **12**:38.

Ramsing, N. B., M. J. Ferris and D. M. Ward (2000). "Highly ordered vertical structure of *Synechococcus* populations within the one-millimeter thick photic zone of a hot spring cyanobacterial mat." *Applied and Environmental Microbiology* **66**:1038-1049.

Reche, I., E. Pulido-Villena, R. Morales-Baquero and E. O. Casamayor (2005). "Does ecosystem size determine aquatic bacterial richness?" *Ecology* **86**:1715-1722.

Reeder J. and R. Knight (2009). "The ′rare biosphere′: a reality check." *Nature Methods* **6**:636-637.

Reno, M. L., N. L. Held, C. J. Field, P. V. Burke and R. J. Whitaker (2009). "Biogeography of the *Sulfolobus islandicus* pan-genome." *Proceedings of the National Academy of Science* **106**:8605-8610.

Rhoads, D. D., R. D. Wolcott, Y. Sun and S. E. Dowd (2012). "Comparison of culture and molecular identification of bacteria in chronic wounds." *International Journal of Molecular Sciences* **13**:2535-2550.

Rocap, G, D. L. Distel, J. B. Waterbury and S. W. Chisholm (2002). "Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences." *Applied and Environmental Microbiology* **68**:1180-1191.

Romme, W. H., M. S. Boyce, R. Gresswell, E. H. Merrill, G. W. Minshall, C. Whitlock and M. G. Turner (2011). "Twenty Years After the 1988 Yellowstone Fires: Lessons About Disturbance and Ecosystems." *Ecosystems* **14**:1196-1215.

Roselló-Mora, R. and R. Amann (2001). "The species concept for prokaryotes." *FEMS Microbiology Review* **25**:39-67.

Ruff-Roberts, A. L., J. G. Kuenen and D. M. Ward (1994). "Distribution of cultivated and uncultivated cyanobacteria and *Chloroflexus*-like bacteria in hot spring microbial mats." *Applied and Environmental Microbiology* **60**:697-704.

Sabehi, G., B. C. Kirkup, M. Rozenberg, N. Stambler, M. F. Polz and O. Béja (2007). "Adaptation and spectral tuning in divergent marine proteorhodopsins from the eastern Mediterranean and the Sargasso Seas." *ISME Journal* **1**:48-55.

Shapiro B. J., J. Friedman, O. X. Cordero, S. P. Preheim, S. C. Timberlake, G. Szabó, M. F. Polz, and E. J. Alm (2012). "Population genomics of early events in the ecological differentiation of bacteria." *Science* **336**:48-51.

Simpson, G. G. (1961). "The species concept." *Evolution* **5**:285-298.

Smith, N. H., S. V. Gordon, R. de la Rua-Domenech, R. S. Clifton-Hadley, and R. G. Hewinson (2006). "Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*." *Nature Reviews Microbiology* **4**:670-681.

Soyer, Y., R. H. Orsi, L. D. Rodriguez-Rivera, Q. Sun and M. Wiedmann (2009). "Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected *Salmonella* serotypes." *BMC Evolution Biology* **9**:264.

Stackenbrandt, E. and J. Ebers (2006). "Taxonomic parameters revisited: tarnished gold standards. Microbiology Today" **33**:152-155.

Stackenbrandt, E. and B. M. Geobel (1994). "Taxonomic note: a place for DNA: DNA reassociation and the 16S rRNA sequence analysis in the present species definition in bacteriology." *International Journal of Systematic Bacteriology* **44**: 846-849.

Steunou, A-S., S. I. Jensen, E. Brecht, E. D. Becraft, M. M. Bateson, O. Kilian, D. Bhaya, D. M. Ward, J. W. Peters, A. R. Grossman and M. Kühl (2008). "Regulation of *nif* gene expression and the energetics of N_2 fixation over the diel cycle in a hot spring microbial mat." *ISME Journal* **2**:364-378.

Tamura K, J. Dudley, M. Nei and S. Kumar (2007). "MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0." *Molecular Biology and Evolution* **24**:1596-1599.

Tamura, K, D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar (2011). "MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods." *Molecular Biology and Evolution* **28**:2731-2739.

Taylor, M. W., P. J. Schupp, R. De Nys, S. Kjelleberg and P. D. Steinberg (2005). "Biogeography of bacteria associated with the marine sponge *Cymbastela concentrica*." *Environmental Microbiology* **7**:419-433.

Thompson, J. R., M. A. Randa, L. A. Marcelino, A. Tomita-Mitchell, E. Lim and M. F. Polz (2004). "Diversity and dynamics of a North Atlantic Coastal *Vibrio* community." *Applied and Environmental Microbiology* **70**:4103-4110.

Torres, P. A., A. B. Abril and E. H. Bucher (2005). "Microbial succession in litter decomposition in the semi-arid Chaco woodland." *Soil Biology and Biochemistry* **37**:49-54.

Torsvik, V., J. Goksøyr and F. L. Daae (1990). "High diversity in DNA of soil bacteria." *Applied and Environmental Microbiology* **56**:782-787.

Van Valen, L. (1976). "Ecological species, multispecies, and oaks." *Taxonomy* **25**:233-239.

van Winkelhof, A. J., U. van der Velden and J. de Graaff (1988). "Microbial succession in recolonizing deep periodontal pockets after a single course of supra- and subgingival debridement." *Journal of Clinical Periodontology* **15**:116-122.

Vos, M. and X. Didelot (2009). "A comparison of homologous recombination rates in bacteria and archaea." *ISME Journal* **3**:199-208.

Walk, S. T., E. W. Alm, L. M. Calhoun, J. M. Mladonicky and T. S. Whittam (2007). "Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches." *Environmental Microbiology* **9**:2274-2288.

Walk S. T., E. W. Alm, D. M. Gordon, J. L. Ram, G. A. Toranzos, J. M. Tiedje and T. S. Whittam (2009). "Cryptic lineages of the genus *Escherichia*." *Applied and Environmental Microbiology.* **75:**6534-6544.

Ward, D. M. (1998). "A natural species concept for prokaryotes." *Current Opinions in Microbiology* **1**:271-277.

Ward, D.M. (2006). "A macrobiological perspective on microbial species." *Microbe* **1**:269-278.

Ward, D. M., M. M. Bateson, M. J. Ferris, M. Kühl, A. Wieland, A. Koeppel and F. M. Cohan (2006). "Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function." *Philosophical Transactions of the Royal Society of London B* **361**:1997-2008.

Ward, D. M. and R. W. Castenholz (2000). "Cyanobacteria in geothermal habitats." *In* The Ecology of Cyanobacteria. Whitton, B.A. and M. Potts (eds). The Netherlands: Kluwer Academic Publishers, pp. 39-57.

Ward, D. M., R. W. Castenholz and S. R. Miller (2012). "Ecology of Cyanobacteria II: Their diversity in Space and Time." *In* Cyanobacteria in geothermal habitats. B. A. Whitton (ed). Springer Science and Business media B. V., pp. 39-63.

Ward, D. M., F. M. Cohan, D. Bhaya, J. F. Heidelberg, M. Kuhl and A. Grossman (2008). "Genomics, environmental genomics and the issue of microbial species." *Heredity* **100**: 207-219.

Ward, D. M. and F. M. Cohan (2005). "Microbial diversity in hot spring cyanobacterial mats: pattern and prediction. *In* Geothermal Biology and Geochemistry in Yellowstone National Park. Inskeep W. P. and T. McDermott (eds.). Thermal Biology Institute: Bozeman, MT, pp 185-201.

Ward, D. M., M. J. Ferris, S. C. Nold and M. M. Bateson (1998). "A natural view of microbial biodiversity within hot spring cyanobacterial mat communities." *Microbiology and Molecular Biology Reviews* **62**:1353-1370.

Ward, D. M, R. Weller and M. M. Bateson (1990). "16S rRNA sequences reveal numerous uncultured microorganisms in a natural community." *Nature* **344**:63-65.

Ward, D. M., R. Weller, J. Shiea, R. W. Castenholz and Y. Cohen (1989). "Hot spring microbial mats: Anoxygenic and oxygenic mats of possible evolutionary significance." *In* Microbial Mats: Physiological ecology of benthic microbial communities. Y. Cohen and E. Rosenberg (eds.). American Society of Microbiology, Washington DC, pp. 3-15.

Wayne, L. G, D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, W. E. C. Moore, R. G. E. Murray, E. Stackenbrandt, M. P. Starr and H. G. Trüper (1987). "Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics." *International Journal Systematic Bacteriology* **37**:463-464.

West, N. J. and D. J. Scanlan (1999). "Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean." *Applied and Environmental Microbiology* **65**:2585-2591.

Whitaker, R. J., D. W. Grogan and J. W. Taylor (2003). "Geographic barriers isolate endemic populations of hyperthermophilic Archaea." *Science* **301**:976-978.

Woese, C. R. (1987). "Bacterial evolution." *Microbiological Reviews* **51**:221-271.

Woese, C. R., O. Kandler and M. L. Wheels (1990). "Towards a natural system of organisms: proposal for the domains of Archaea, Bacteria, and Eucarya." *Proceedings of the National Academy of Sciences* **87**:4576-4579.

Xavier, D., D. Lawson, A. Darling and D. Falush (2010). "Inference of homologous recombination in bacteria using whole-genome sequences." *The Genetics Society of America* **186**:1435-1449.

Zhou, D., Y. Han, Y. Song, Z. Tong, J. Wang, Z. Guo, D. Pei, X. Pang, J. Zhai, M. Li, B. Cui, Z. Qi, L. Jin, R. Dai, Z. Du, J. Bao, X. Zhang, J. Yu, J. Wang, P. Huang and R. Yang (2004). "DNA microarray analysis of genome dynamics in *Yersinia pestis*: insights into bacterial genome microevolution and niche adaptation." *Journal of Bacteriology* **186:**5138-5146.

APPENDICES

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

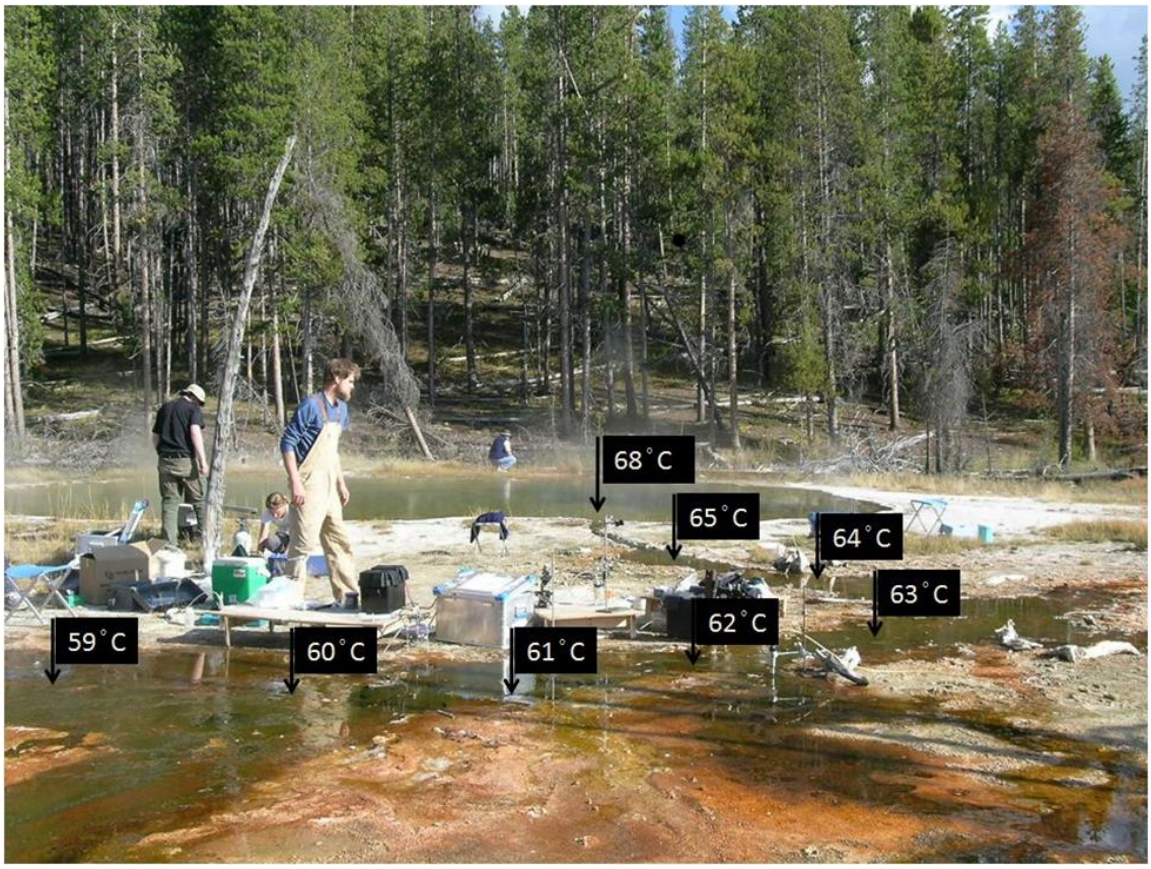


Figure A.1. Approximate locations of samples collected at Mushroom Spring on 13 and 14 September 2008, and the temperatures measured at time of sampling. Samples collected were used for denaturing gradient gel electrophoresis analysis shown in Appendix A Figure A3.

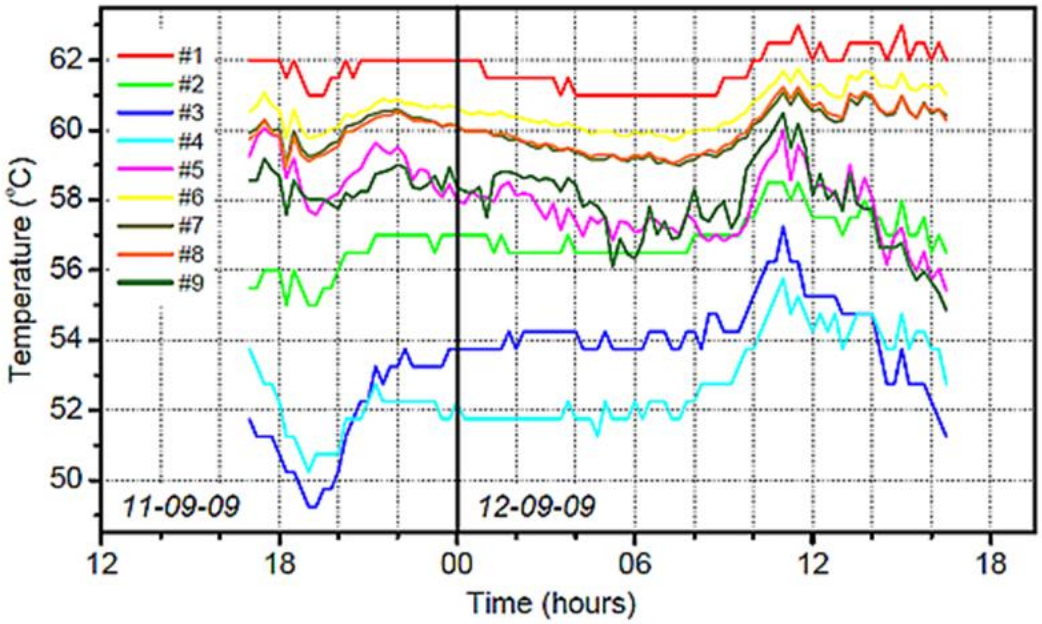


Figure A.2. Daily temperature fluctuations a 9 sites in Mushroom Spring main flow and side effluent channel sites over a diel cycle. Measurements were taken on 11 and 12 September 2009. These recordings were made using the DS1922L/T temperature logger iButtons (Maxim, Sunnyvale, CA, USA).

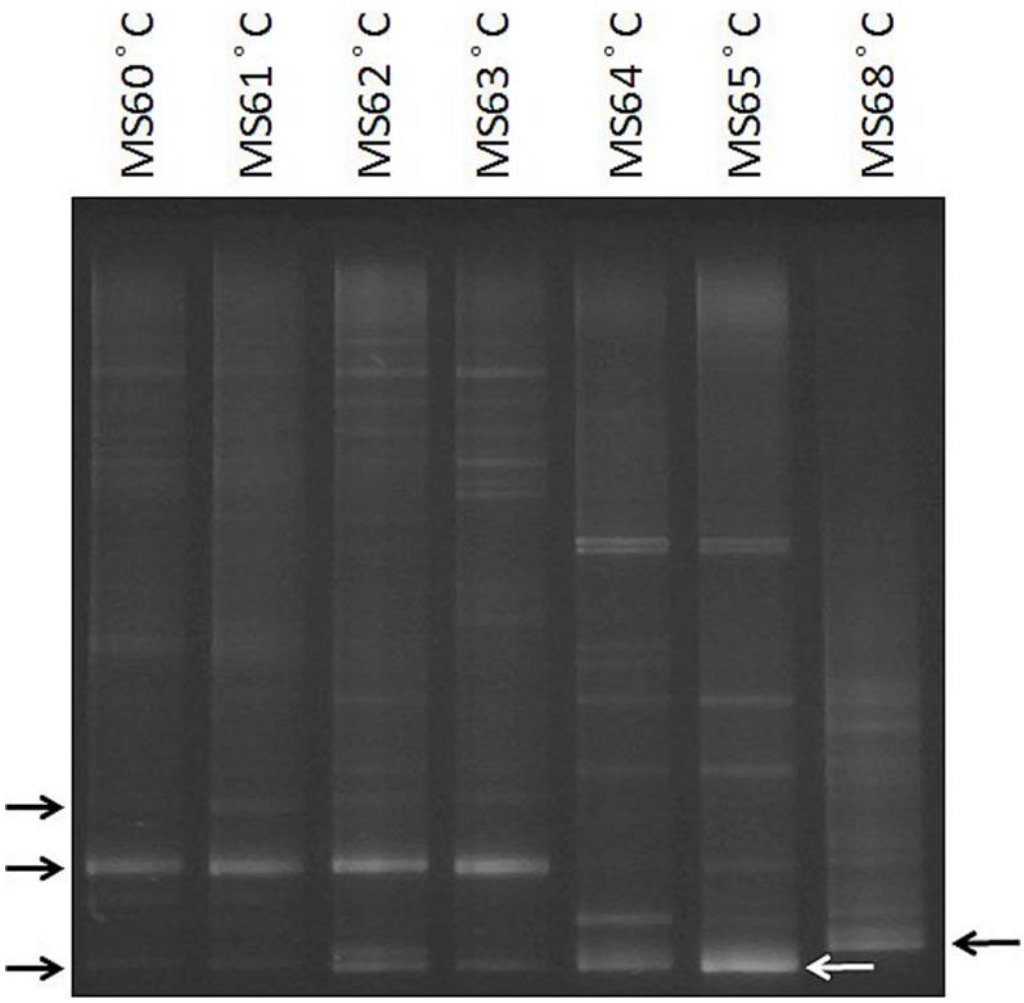


Figure A.3. Denaturing gradient gel electrophoresis analysis of *psaA* sequences PCR-amplified from DNA extracted from mat samples collected on 13 and 14 September 2008 along the main flow path of Mushroom Spring, Yellowstone National Park. Samples (see Appendix A Figure A1). Arrows show correspondence with bands in main-text Figure 3 in terms of relative position, however, exact positions differ due to gel to gel variation. Note that in this collection the 65°C sample resembles the 64°C sample, whereas, the 68°C sample resembles the 65°C sample shown in main-text Figure 3.

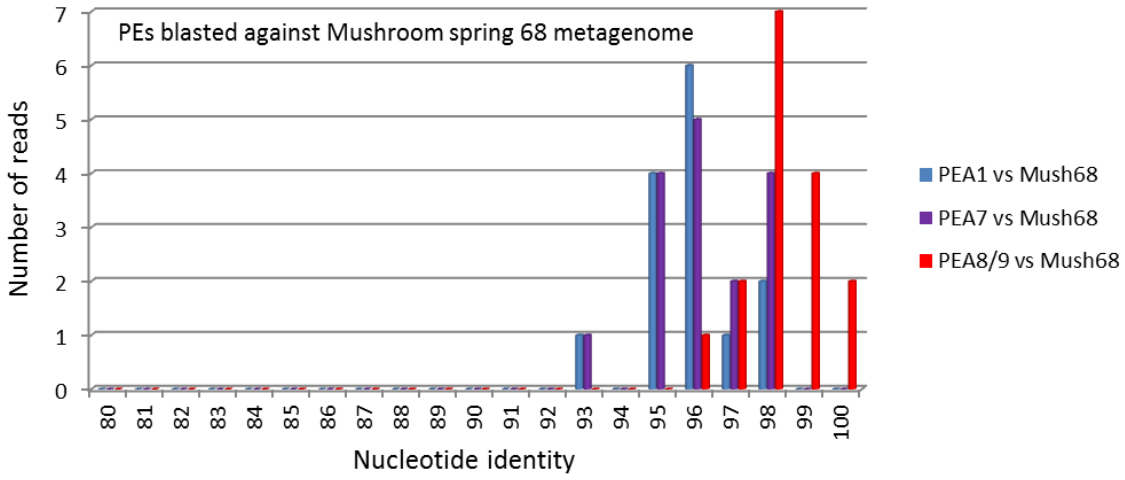


Figure A.4. Sequence similarity of *psaA* temperature-defined PEs A1, A7, A8 and A9 compared to *psaA* variants in the Mushroom Spring 68°C metagenome.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

Section I: Comparison of Putative Ecotypes (PEs) Demarcated From High-frequency *psaA* Sequences by Ecotype Simulation Using the Fine-Scale or Conservative Demarcation Approaches

The conservative approach allowed for the demarcation of the more divergent B′-like populations, where all sequence variants within a given predicted ecotype shared the same environmental distribution (main-text Figure 1B). However, the conservative approach grouped A-like phylogenetic groups with different distributions and responses to environmental perturbations as a single ecotype (e.g. PEs A6 and A12 are lumped into one PE) (Appendix Figure B2 A). In contrast, the fine-scale approach consistently demarcated A-like PEs in accord with their distributions (main-text Figure 1A), while it split B′-like populations with near-identical distributions that could not be distinguished ecologically into separate ecotypes (e.g. PE B′8 subclades were split into three separate PEs) (Appendix Figure B2 B).

Section II: Additional Vertical *psaA* Profiles and Temperature Variations in the Mat.

Spatio-temporal temperature variation complicated exact replication in the vertical distribution studies. Reproducibility among duplicate cores for vertical profiling at different temperatures was evident in pattern, but differences in the relative abundances of PEs suggested that natural variation, sequencing error or temperature effects were responsible for sample variation. Samples were not always collected from exactly the same temperature, though patterns of vertical PE distribution remained consistent, indicating that temperature effects were the likely cause for the observed variation between samples. This was likely due to temperature variations that occurred

perpendicular to flow. In the main text, we therefore show an individual vertical profile as an example pattern of PE distribution at a site thought to be representative of EZ1, and we show all additional vertical profiles in Appendix B Figure B3. EZs for each vertically dissected core were determined after summing the numbers of variants in each PE in all layers of a core, and comparison to the relative abundances of PEs in whole cores used to generate the temperature profile in main-text Figure 3.2 (Appendix B Table B5).

Vertical profiles of predominant PEs in samples ostensibly collected at average midday temperature sites of ~60, 63 and 65°C corresponded to EZ1 PE diversity (60 and 63°C flow path samples) (Appendix B Figure B3 A-D; see Appendix B Tables B6-11 for vertical distributions of less abundant PEs). Vertical profiles of samples collected at 60 and 63°C appear to be from slightly cooler temperatures in EZ1, where there is less total A-like diversity and PE B′2 is higher in relative abundance, possibly indicating within-EZ fine-scale temperature adaptation. The 65°C vertical profile (Appendix B Figure B3 D), which was targeted for EZ2, appears to be a near-replicate of main-text Figure 3A (EZ1), indicating that it was taken at a slightly cooler temperature perpendicular to the main flow path (Appendix B Figure B8).

The 65°C profile that is representative of EZ2 was comprised of only A-like PEs, with PEs A1, A7 and A12 predominant in surface layers and PEs A6 and A14 maximizing 400-480 µm below the mat surface (Appendix B Figure B3 E). PEs A1, A7 and A12 were also abundant in the deepest layers, raising concern that the short *psaA* sequences used for PE demarcation may not have contained sufficient molecular variation to resolve extremely closely related ecologically distinct A-like populations in

these cases (e.g. PE A1; see main-text Methods). All PEs in all EZs had significantly different distributions compared to other predominant PEs along the vertical aspect of the mat ($P<0.001$; Appendix B Table B1).

Section III: *psaA* Expression Patterns in Ecological Zone (EZ) 1, EZ2 and EZ3 During Morning and Evening Light Transition Periods

In addition to the full diel *psaA* expression pattern done on samples typical of EZ1 (main-text Figure 4), *psaA* transcripts were also investigated at selected time points at sites representative of EZ1, EZ2 and EZ3 in terms of the PE transcripts present.

Methods. Samples for partial diel expression analysis were collected at time points corresponding to low-light anoxic (early morning and late evening), low-light oxic (late morning), and high-light oxic (mid-day) conditions to focus on transition periods when ecotype-specific expression might be observed. These samples were collected with a #4 cork borer (38.5 mm^2) at the average midday temperature of 63°C on 12 and 13 September 2008, at 65°C on 13 and 14 September 2008, and at 68°C on 12 and 13 September 2009. Samples were collected from each site at 0500, 0730, 0830, 0930, 1100, 1300, 1830 and 1930 h. Some time points were lost due to failed sequencing reactions. Extraction and purification of RNA, cDNA synthesis and data analysis methods were the same as described in the main-text Materials and Methods. The 22 samples analyzed for partial diel expression yielded a total of 7790 sequences (average of 363; standard deviation of 122). Sequences were trimmed to 247 base pairs to obtain the maximum number of sequences, cleaned and analyzed to identify high-frequency sequences

(sequences equaling ≥50 identical copies in all samples) and associated low frequency sequences (sequences equaling <50 copies in all samples), and PEs were demarcated as described in the main-text.

Results and Discussion. Patterns of gene expression differed for all predominant PEs; results are shown in Appendix B Figure B5 ($P<0.001$; Appendix B Table B2). At the site representative of EZ1, the expression patterns of the most abundant PEs resembled those observed in the full diel study (main-text Figure 4), except that PE B′9 transcripts declined and PE A1 transcripts peaked earlier in the day. This could have been due to heterogeneity in thickness of upper green layers, leading to deeper light penetration into the mat earlier in the day causing a rise in PE A1 transcripts at an earlier time point. Among less abundant PEs, transcripts of the subsurface PE A6 (main-text Figure 3A) were highest between 0930 and 1100 h, when transcripts of PE A1, also a subsurface population, were highest. Differences in transcript abundance in the two experiments may also have resulted from pooling samples over a ~1 m^2 area in the full diel study, as opposed to studying replicate samples collected over a ~30 cm^2 area in partial diel studies.

At the site representative of EZ2, transcripts of PE A1 predominated at all time points and reached a peak at 1100 h under a photon irradiance (400-700 nm) of ~1500 μmol photons m^{-2} s^{-1}. A possible explanation for the absence of a mid-day decrease in transcripts is that PE A1 may be comprised of more than one extremely closely related ecotype population with different depth distributions (see main-text Materials and Methods). Although PE A1 was most abundant near the mat surface, there was a second

peak in its abundance in deeper mat layers (Appendix B Figure B3 E), supporting this possibility. Transcripts of the subsurface PE A6 peaked in the early morning (0830 h; 50-200 µmol photons m^{-2} s^{-1}) and early evening (1830 h; 200-400 µmol photons m^{-2} s^{-1}). A possible reason for the earlier timing of expression in this PE is its higher positioning in the mat in EZ2 compared to EZ1. Transcripts of PE A7, which is evenly distributed throughout the entire mat (Appendix B Figure B3 E), peaked in relative abundance at 0730 (1-50 µmol photons m^{-2} s^{-1}) and 1100 h (~1400 µmol photons m^{-2} s^{-1}), but maintained fairly constant transcript levels throughout the day until 1700 h.

At the site representative of EZ3, PE A′8 transcripts peaked at 1100 h under solar photon irradiance (400-700 nm) of >1600 µmol photons m^{-2} s^{-1}, reaching nearly 70% of the total *psaA* relative transcript abundance. PE A′9 transcripts exhibited a mid-day minimum and peaked in the evening (1830 h; 200-400 µmol photons m^{-2} s^{-1}). Where error bars are missing, replication was not possible due to failed sequencing reactions, but the available data suggested that the transcript patterns of the two predominant PEs were different. Fine-scale vertical dissection of samples from EZ3 was not possible due to the loose density of the mat at higher temperatures, although differential distribution of *Synechococcus* populations was noted at this temperature by Ferris et al. (2003) (see also Ward et al, 2006). In these studies clones containing 16S-23S rRNA internal transcribed spacer region sequences, which were prepared from samples obtained using a syringe and needle to collect from the upper and lower parts of the mat green layer, were examined. This suggested the presence of vertical structure in mats in EZ3.

Section IV. Controls and Replicated
UV-light Removal Experiments.

Samples from and experiment performed in 1996 were analyzed to determine if PE relative abundances changed over time at undisturbed control sites, and to replicate the effects of UV blocking on *Synechococcus* mat populations.

Methods. A 58.6 to 62.2°C (EZ1) site in the microbial mat of MS was covered with a wire mesh screen supported by a wooden platform placed ~1 to 2 cm above the water surface to protect the experimental area from possible hail damage. UV was blocked as described in main-text Materials and Methods. Samples were collected in duplicate by David Ward from 21 to 26 July 1996 using a #4 cork borer (38.5 mm^2) every day between 1200 and 1800 h. Samples were immediately frozen and stored at -80°C until analysis in 2012. Molecular methods were the same as described in the main-text.

Results and Discussion. PE diversity of the undisturbed experimental site mostly matched the diversity observed in EZ1, where B′-like PEs are the predominant populations. In the unaltered control samples there was minimal change in the relative abundances of predominant PEs over a 5 day period (P = 0.35; Appendix B Figure B6 A and Table B2).

After blocking UV-light, PE B′9 increased in relative abundance by more than 5-fold 1 day after light alteration, despite being in low-abundance initially in 1996 in the mat at this site (i.e., <1% contribution to the *psaA* pool), and declined in relative abundance after day 2. PEs B′8 and B′12 increased in relative abundance to a lesser extent after UV exclusion, and returned to their approximate Day 0 relative abundances

by day 5. While not all predominant populations were the same when comparing PEs demarcated in 1996 to PEs demarcated in 2008, the initial increase of PE B′9 after UV blocking in 1996 was similar to what was observed in the 2008 Ti454-barcoding study (P=0.03; Appendix B Figure B6 B and Table B3).

Table B1. Summary of G-test statistical analyses for flow path, vertical distributions, 2008 UV exclusion experiment and PE-specific transcription studies. Temperature and ecological zones (EZs) are indicated for each sample. Statistics for main-text distributions are in bold. Letters in paraentheses correspond to Appendix B Figure B3 (vertical) and 5 (transcription).

Distribution	Number of samples analyzed	Number of ecotypes analyzed	Degrees of freedom	chi square value	P-value
flow path A/B′	**4**	**11**	**30**	**1.10E+04**	**<0.001**
flow path A	**4**	**6**	**15**	**4.20E+03**	**<0.001**
flow path B′	**4**	**5**	**12**	**4.99E+02**	**<0.001**
vertical 63°C EZ1 A/B′	**10**	**7**	**54**	**4.30E+03**	**<0.001**
vertical 63°C EZ1 A	**10**	**4**	**27**	**6.72E+02**	**<0.001**
vertical 63°C EZ1 B′	**10**	**3**	**18**	**1.30E+02**	**<0.001**
vertical 60°C EZ1 A/B′ (3A)	5	6	20	6.71E+02	<0.001
vertical 60°C EZ1 A (3A)	5	3	8	1.23E+02	<0.001
vertical 60°C EZ1 B′ (3A)	5	3	8	1.04E+02	<0.001
vertical 60°C EZ1 A/B′ (3B)	10	6	45	1.50E+03	<0.001
vertical 60°C EZ1 A (3B)	10	3	18	1.13E+02	<0.001
vertical 60°C EZ1 B′ (3B)	10	3	18	8.13E+02	<0.001
vertical 63°C EZ1 A/B′ (3C)	10	4	27	2.00E+03	<0.001
vertical 63°C EZ1 A (3C)	10	2	18	5.97E+01	<0.001
vertical 63°C EZ1 B′ (3C)	10	2	18	9.56E+02	<0.001
vertical 65°C EZ1 A/B′ (3D)	12	6	55	6.90E+03	<0.001
vertical 65°C EZ1 A (3D)	12	4	33	2.90E+03	<0.001
vertical 65°C EZ1 B′ (3D)	12	2	11	1.46E+02	<0.001
vertical 65°C EZ2 A (3E)	9	6	40	7.82E+02	<0.001
transcription 60°C EZ1	**18**	**4**	**48**	**6.93E+02**	**<0.001**
transcription 63°C EZ1(5A)	4	3	6	1.69E+02	<0.001
transcription 65°C EZ2 (5B)	5	3	8	0.89E+02	<0.001
transcription 68°C EZ3 (5C)	4	2	3	0.31E+02	<0.001
UV exclusion	4	4	6	0.54E+02	<0.001

Table B2. Summary of Ancova statistical analyses for perturbation studies and controls. Statistics for main-text distributions are in bold.

Experiment (year)	P-value	Fixation indices	Degrees of freedom
Control (1996)	0.35	1.15	2,24
UV exclusion (1996)	0.03	0.91	3,12
Light reduction (2008)	**0.026**	**4.40**	**3,12**
Temperature shift (2008)	**0.016**	**3.83**	**5,18**

Table B3. p–values associated with the hypothesis that the ratio of the number of singleton and low-frequency multi-sequence variants (LFSs) to the number of HFSs in predominant PE clades were different over the four day light reduction experiment (A) and temperature shift experiment (B).

(A)

PE	p-value
A1	0.3423
A4	0.7679
A6	0.8335
A12	0.5236

(B)

PE	p-value
B′2	0.2506
B′8	0.4586
B′9	0.5062
A1	0.5236
A6	0.5453
A14	0.9529

Table B4. PE population percentage along replicate effluent flow path samples.

	Replicate sample 1				Replicate sample 2			
PE	**60°C**	**63°C**	**65°C**	**68°C**	**60°C**	**63°C**	**65°C**	**68°C**
A1	0.8	1.5	22.5	7.3	4.3	4.2	31.0	2.0
A2	0.3	0.0	0.2	3.6	0.0	0.6	0.5	0.3
A3	0.1	0.1	0.6	0.0	0.1	0.0	0.3	0.0
A4	0.1	0.4	5.2	0.8	5.2	0.5	2.0	0.3
A5	0.2	0.4	4.6	1.0	2.7	1.6	11.6	0.2
A6	0.3	1.9	16.2	2.9	12.5	4.5	10.2	0.9
A7	0.0	0.1	0.1	4.9	0.0	0.2	8.9	1.4
A′8	0.0	0.0	0.0	7.9	0.0	0.0	0.0	13.4
A′9	0.0	0.0	0.2	17.0	0.0	0.0	0.3	29.1
A10	0.2	0.0	0.9	0.8	0.4	0.3	0.6	0.1
A11	0.0	0.0	2.2	0.1	0.2	0.1	0.5	0.0
A12	0.0	0.5	4.0	2.5	0.1	0.1	21.2	1.1
A13	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3
A14	0.1	0.5	12.6	1.1	3.8	2.0	8.7	0.0
A15	0.0	0.0	0.7	0.1	0.1	0.0	0.6	0.0
A′16	0.0	0.0	0.1	6.2	0.0	0.1	0.3	10.4
A′17	0.0	0.0	0.0	10.3	0.0	0.0	0.0	18.9
A′18	0.0	0.0	0.0	0.4	0.0	0.0	2.5	0.3
A19	0.0	0.0	0.0	3.7	0.0	0.0	0.0	5.9
A20	0.0	0.0	0.1	1.4	0.0	0.1	0.1	2.2
A′21	0.0	0.0	0.2	6.4	0.0	0.1	0.2	9.4
A′22	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.1
B′1	0.5	0.3	0.0	0.4	0.3	0.2	0.0	0.3
B′2	21.3	13.5	2.6	1.0	10.0	5.8	0.1	0.0
B′3	0.0	0.3	0.0	1.7	0.2	0.3	0.0	0.0
B′4	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0
B′5	0.0	0.3	0.1	0.0	0.0	0.2	0.0	0.0
B′6	0.1	0.0	0.0	1.4	0.0	0.0	0.0	0.1
B′7	1.1	7.0	0.1	0.3	0.4	3.3	0.0	0.1
B′8	8.8	2.2	2.6	1.4	8.3	2.7	0.6	0.5
B′9	23.5	29.0	6.1	4.7	23.4	26.8	0.7	1.6
B′10	0.0	0.0	0.3	0.0	0.3	0.0	0.0	0.0
B′11	0.6	0.6	0.2	0.0	0.6	1.3	0.0	0.1
B′12	5.9	9.3	2.0	1.6	3.6	6.2	0.0	0.6
B′13	0.1	0.1	0.0	0.1	0.1	0.2	0.0	0.0
B′14	0.0	1.5	0.0	0.0	0.0	0.5	0.0	0.1
B′15	0.0	0.4	6.8	1.2	0.4	0.1	3.3	0.8
B′16	0.0	0.7	0.0	0.0	0.0	0.2	0.0	0.0
B′17	0.8	0.2	0.0	0.0	0.9	1.3	0.0	0.0
B′18	0.9	0.8	0.3	0.1	0.7	0.5	0.0	0.0
B′19	0.1	0.8	0.0	0.0	0.2	0.5	0.0	0.0
B′20	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0
B′21	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0
B′22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
B′23	0.1	0.7	0.0	0.0	0.2	0.5	0.0	0.0
B′24	0.0	0.0	0.2	0.1	0.1	0.0	0.0	0.0

Table B5. Percent population of relatively abundant A-like, A′-like and B′-like PEs in pooled vertically dissected samples in relation to temperature samples used in Chapter 3, Figure 3.2. Shaded columns indicate EZ of samples (blue, EZ1; purple, EZ2; red, EZ3).

	60°C	60°C	60°C Vertical 1	60°C Vertical 2	63°C	63°C	63°C Vertical 1	63°C Vertical 2	65°C	65°C	65°C Vertical 1	65°C Vertical 2	68°C	68°C
A1	0.8	4.3	6.6	6.7	1.5	4.2	6.7	15.9	22.5	31	14.4	29.1	7.3	2
A4	0.1	5.2	2.6	1.5	0.4	0.5	1.5	7.8	5.2	2	4.6	2.8	0.8	0.3
A5	0.2	2.7	1.2	0.9	0.4	1.6	0.9	2.9	4.6	11.6	2.4	5.2	1	0.2
A6	0.3	12.5	6.4	2.9	1.9	4.5	2.9	9.3	16.2	10.2	15.2	12.5	2.9	0.9
A7	0	0	0.2	0.1	0.1	0.2	0.1	1.6	0.1	8.9	0.4	10.9	4.9	1.4
A′8	0	0	0	0	0	0	0	0	0	0	0	0	7.9	13.4
A′9	0	0	0	0.1	0	0	0.1	0.2	0.2	0.3	0.2	0.4	17	29.1
A12	0	0.1	0.1	0.2	0.5	0.1	0.2	0.9	4	21.2	2	23.5	2.5	1.1
A14	0.1	3.8	4.8	4.2	0.5	2	4.2	5.3	12.6	8.7	16.3	5.5	1.1	0
A′16	0	0	0	0.1	0	0.1	0.1	0	0.1	0.3	0.1	0.3	6.2	10.4
A′17	0	0	0	0	0	0	0	0	0	0	0.2	0.4	10.3	18.9
B′2	21.3	10	14.4	25.8	13.5	5.8	25.8	1.8	2.6	0.1	3.6	0	1	0
B′7	1.1	0.4	0.8	0.7	7	3.3	0.7	3.8	0.1	0	0.4	0	0.3	0.1
B′8	8.8	8.3	6.8	9.2	2.2	2.7	9.2	3.1	2.6	0.6	2.2	0.6	1.4	0.5
B′9	23.5	23.4	23.9	21	29	26.8	21	16.2	6.1	0.7	11.6	2.3	4.7	1.6
B′12	5.9	3.6	3.1	4.9	9.3	6.2	4.9	2.1	2	0	3	0.4	1.6	0.6
B′15	0	0.4	0.1	0.3	0.4	0.1	0.3	7	6.8	3.3	1.7	1.4	1.2	0.8

Table B6. PE population percentages along the vertical gradient at the ~63°C average midday temperature site (EZ1; main-text Figure 3A).

	% PE along 80 µm depth intervals Depth interval									
PE	**0/80 µm**	**80/160 µm**	**160/240 µm**	**240/320 µm**	**320/400 µm**	**400/480 µm**	**480/560 µm**	**560/640 µm**	**640/720 µm**	**720/800 µm**
A1	7.7	11.5	11.2	12.2	16.4	20.3	20.3	21.1	19.5	16.7
A2	0.0	0.0	0.2	0.2	0.3	0.2	0.2	0.7	0.3	0.6
A3	4.0	1.5	1.1	2.0	0.6	0.4	0.4	0.1	0.0	0.3
A4	0.2	5.2	3.3	8.1	7.7	8.9	8.1	8.8	16.0	11.8
A5	1.2	2.6	3.5	1.4	1.4	1.1	5.0	3.3	3.0	7.2
A6	0.1	4.3	3.1	5.6	5.8	5.9	7.2	12.1	24.3	25.6
A7	0.3	1.2	1.4	0.2	1.1	2.6	2.2	3.2	1.8	1.5
A′8	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
A′9	0.5	0.3	0.4	0.1	0.3	0.1	0.0	0.0	0.1	0.0
A10	0.6	0.6	0.8	0.5	0.4	0.2	0.1	0.0	0.1	0.1
A11	0.0	0.2	0.6	0.3	0.6	1.1	0.9	2.0	4.4	5.2
A12	0.1	0.7	1.2	1.7	1.2	0.7	0.9	1.0	0.6	0.5
A13	0.0	0.0	0.0	0.1	0.2	0.1	0.1	0.0	0.1	0.3
A14	0.9	3.2	2.4	3.0	3.0	4.3	4.5	7.4	13.2	12.2
A15	0.0	0.0	0.2	0.1	0.1	0.0	0.4	0.7	0.1	0.1
A′16	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.0
A′17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A′18	0.2	0.2	0.2	0.0	0.3	0.0	0.5	0.7	0.3	0.3
A19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A20	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
A′21	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A′22	1.8	0.4	0.7	0.3	0.4	0.3	0.2	0.0	0.0	0.0
B′1	0.1	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.2
B′2	4.5	1.6	2.1	1.7	2.0	1.8	1.9	0.8	0.5	0.8
B′3	0.2	0.5	0.4	0.5	0.3	0.4	0.1	1.2	0.4	0.2
B′4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
B′5	0.1	0.5	0.6	0.9	0.4	1.0	0.4	0.8	0.4	0.4

B′6	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.1	0.0	0.0
B′7	11.3	6.6	6.1	4.4	3.0	2.5	2.8	1.3	0.3	0.1
B′8	1.6	2.7	3.9	2.7	4.0	4.8	3.3	4.0	1.1	1.9
B′9	35.0	19.8	19.7	23.0	21.3	16.8	13.3	7.6	2.1	2.3
B′10	0.1	0.3	0.2	0.7	0.2	0.3	0.6	0.1	0.1	0.0
B′11	0.9	0.1	0.4	0.2	0.5	0.0	0.5	0.0	0.0	0.0
B′12	7.1	9.2	7.2	7.2	5.9	6.6	6.3	5.4	1.9	2.5
B′13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′14	0.2	0.2	0.2	0.3	0.5	0.2	0.4	0.7	0.2	0.2
B′15	10.4	12.2	10.1	10.1	8.0	5.9	7.1	4.6	1.2	0.9
B′16	0.4	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
B′17	0.3	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.2	0.1
B′18	0.7	1.2	1.2	0.1	0.1	0.3	0.2	0.0	0.0	0.0
B′19	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
B′20	0.0	0.0	0.0	0.0	0.1	0.2	0.2	0.0	0.0	0.0
B′21	0.0	0.1	0.3	0.0	0.0	0.0	0.0	0.2	0.0	0.0
B′22	0.3	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.1
B′23	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′24	0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table B7. PE population percentages along the vertical gradient at the ~60°C average midday temperature site (EZ1; Appendix B Figure B3 A).

	% PE along 80 µm depth intervals									
PE	**0/80 µm**	**80/160 µm**	**160/240 µm**	**240/320 µm**	**320/400 µm**	**400/480 µm**	**480/560 µm**	**560/640 µm**	**640/720 µm**	**720/800 µm**
A1	1.8	3.3	6.0	6.8	6.0	7.7	8.9	9.3	8.9	1.8
A2	0.0	0.1	0.4	0.0	0.6	0.2	0.1	0.0	0.2	0.0
A3	0.2	0.1	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.2
A4	0.6	0.4	1.7	1.7	2.3	1.8	1.5	2.2	1.9	0.6
A5	0.0	0.5	1.0	0.3	1.4	0.9	1.1	1.7	1.4	0.0
A6	0.2	0.4	2.0	1.4	2.6	3.7	4.1	3.7	6.7	0.2
A7	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
A′8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A′9	0.0	0.1	0.0	0.1	0.2	0.1	0.1	0.1	0.0	0.0
A10	0.1	0.0	0.5	0.0	0.0	0.1	0.0	0.0	0.2	0.1
A11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A12	0.0	0.0	0.9	0.0	0.1	0.1	0.0	0.4	0.0	0.0
A13	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0
A14	0.3	1.6	2.5	2.3	4.4	4.1	3.7	9.7	6.6	0.3
A15	0.0	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.0
A′16	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.5	0.0
A′17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A′18	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
A19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.5	0.0
A′21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A′22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′1	0.2	0.1	0.4	0.0	0.5	1.0	0.3	0.1	0.2	0.2
B′2	18.0	20.8	23.7	21.6	21.2	23.3	27.2	31.8	33.7	18.0
B′3	0.6	2.2	0.5	0.9	1.1	0.7	0.6	0.2	0.4	0.6
B′4	0.1	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.1

B′5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′7	1.6	1.1	0.5	0.9	0.2	0.7	0.7	0.3	0.4	1.6
B′8	8.1	11.6	12.2	12.2	9.2	12.7	10.3	6.8	5.2	8.1
B′9	37.8	29.2	20.5	24.5	25.0	17.7	16.9	9.3	11.9	37.8
B′10	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0
B′11	1.4	0.8	0.8	0.4	0.7	2.6	0.3	0.1	0.3	1.4
B′12	8.9	8.4	5.8	5.7	4.1	3.5	1.9	2.2	3.0	8.9
B′13	0.4	0.2	0.2	0.2	0.1	0.0	0.3	0.1	0.1	0.4
B′14	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.0	0.0
B′15	0.2	0.6	0.3	0.7	0.3	0.2	0.0	0.0	0.0	0.2
B′16	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1
B′17	0.1	0.1	0.1	0.3	0.2	0.2	0.3	0.4	0.5	0.1
B′18	2.2	1.5	1.0	1.0	1.1	1.3	1.5	0.5	0.5	2.2
B′19	0.4	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.4
B′20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′21	0.4	0.3	0.3	0.0	0.1	0.1	0.8	0.3	0.2	0.4
B′22	0.0	0.1	0.4	0.3	0.1	0.2	0.1	0.1	0.1	0.0
B′23	0.2	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2
B′24	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table B8. PE population percentages along the vertical gradient at ~60°C average midday temperature site (EZ1; Appendix B Figure B3 B).

	% PE along 80 µm depth intervals Depth interval				
PE	**0/80 µm**	**80/160 µm**	**160/240 µm**	**240/320 µm**	**320/400 µm**
A1	4.6	7.4	6.7	7.8	7.0
A2	0.2	0.1	0.1	0.1	0.2
A3	0.1	0.1	0.0	0.0	0.0
A4	0.8	2.1	3.5	4.7	3.0
A5	0.8	1.4	0.7	0.8	3.4
A6	0.7	4.1	7.1	11.9	13.6
A7	0.2	0.1	0.3	0.2	0.1
A′8	0.0	0.0	0.0	0.0	0.1
A′9	0.0	0.0	0.0	0.0	0.4
A10	0.1	0.2	0.5	0.7	0.6
A11	0.1	0.8	1.5	1.8	4.5
A12	0.2	0.0	0.0	0.0	0.4
A13	0.1	0.0	0.1	0.1	0.1
A14	0.9	3.2	5.8	8.6	9.4
A15	0.0	0.1	0.5	0.6	0.0
A′16	0.0	0.0	0.0	0.0	0.2
A′17	0.0	0.0	0.0	0.0	0.0
A′18	0.0	0.0	0.0	0.2	0.0
A19	0.0	0.0	0.0	0.0	0.0
A20	0.0	0.0	0.0	0.0	0.0
A′21	0.0	0.0	0.0	0.0	0.0
A′22	0.0	0.0	0.0	0.0	0.1
B′1	0.3	0.3	0.5	0.2	0.2
B′2	13.4	15.0	15.1	15.9	11.9
B′3	2.0	1.5	0.9	1.4	0.7
B′4	0.1	0.0	0.0	0.0	0.0
B′5	0.0	0.1	0.0	0.0	0.0
B′6	0.0	0.0	0.0	0.0	0.2
B′7	1.2	0.4	0.5	1.1	0.8
B′8	7.6	7.6	6.4	6.7	3.9
B′9	37.0	24.6	20.5	14.2	15.0
B′10	0.1	0.0	0.3	0.4	0.6
B′11	1.0	0.3	0.6	0.7	0.0
B′12	3.9	3.6	3.3	2.1	1.8
B′13	2.1	1.3	1.5	0.7	0.1
B′14	0.1	0.0	0.0	0.0	0.0
B′15	0.0	0.2	0.0	0.0	0.1
B′16	0.0	0.0	0.0	0.0	0.0
B′17	0.2	0.5	0.3	0.2	0.4
B′18	2.0	1.1	1.8	0.5	0.1
B′19	0.0	0.5	0.1	0.1	1.1
B′20	0.0	0.3	0.0	0.5	0.0
B′21	0.1	0.1	0.0	0.0	0.0
B′22	0.1	0.0	0.0	0.2	0.0
B′23	0.0	0.0	0.0	0.2	0.0
B′24	0.0	0.0	0.0	0.0	0.0

Table B9. PE population percentages along the vertical gradient at ~63°C average midday temperature site (EZ1; Appendix B Figure B3 C).

	% PE along 80 µm depth intervals Depth interval										
PE	**0/80µm**	**80/160µm**	**160/240µm**	**240/320µm**	**320/400µm**	**400/480µm**	**480/560µm**	**560/640µm**	**640/720µm**	**720/800µm**	**800/880µm**
A1	10.8	13.5	0.8	4.5	8.2	4.7	8.5	1.8	8.4	10.8	1.8
A2	0.2	0.3	0.1	0.0	0.2	0.3	0.2	0.0	0.0	0.0	0.0
A3	0.1	0.2	0.2	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A4	2.2	1.4	0.4	0.3	0.9	2.2	2.1	1.1	3.2	2.9	0.6
A5	2.1	4.1	0.2	1.0	0.8	1.1	2.6	0.4	1.4	2.1	0.0
A6	2.8	2.5	0.0	0.0	0.2	1.4	2.4	1.1	1.7	2.4	1.1
A7	0.0	2.3	0.0	0.2	0.0	0.8	0.3	0.1	0.1	0.0	0.1
A′8	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
A′9	0.0	0.0	0.0	0.0	0.3	0.3	0.0	0.1	0.0	0.0	0.0
A10	0.1	0.8	0.0	0.0	0.3	0.0	0.0	0.0	0.6	0.2	0.0
A11	0.1	0.4	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.0
A12	0.0	1.4	0.0	0.4	0.2	0.5	0.9	0.0	0.0	0.3	0.0
A13	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0
A14	3.2	2.0	0.0	0.5	1.6	1.2	4.0	1.0	2.8	3.6	1.1
A15	0.1	0.2	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0
A′16	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0
A′17	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
A′18	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A′21	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A′22	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
B′1	0.0	0.1	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0
B′2	8.9	3.2	4.6	4.9	10.5	7.5	16.4	6.6	13.9	8.8	4.8
B′3	0.8	1.5	0.1	0.9	1.1	1.6	0.3	1.7	1.3	0.7	0.7
B′4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.2	0.0
B′5	0.1	0.1	0.0	0.0	0.0	0.2	0.1	0.0	0.1	0.0	0.0
B′6	0.0	0.2	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.0
B′7	2.1	2.9	5.8	3.6	2.1	2.6	1.3	3.5	2.2	2.5	11.2

B′8	4.8	2.0	0.4	1.2	5.0	5.8	5.1	2.4	4.7	5.8	0.8
B′9	23.7	40.2	72.8	55.3	35.7	26.1	13.2	48.9	18.9	22.8	50.6
B′10	0.3	0.8	0.0	0.0	0.4	0.3	0.0	0.2	0.0	0.0	0.0
B′11	0.4	0.6	1.3	1.0	1.0	0.5	0.2	0.7	0.5	0.7	1.1
B′12	3.6	5.0	4.7	8.7	3.5	4.4	2.4	7.6	1.8	3.3	7.3
B′13	1.1	0.0	0.0	0.2	2.4	2.8	1.5	5.6	0.8	0.6	4.5
B′14	0.1	0.0	0.1	0.0	0.1	0.7	0.3	0.3	0.2	0.1	1.4
B′15	1.1	1.2	0.0	0.8	0.2	0.5	0.3	0.6	0.1	0.4	0.3
B′16	0.1	0.7	0.8	1.3	0.3	0.1	0.2	0.3	0.4	0.3	0.5
B′17	0.4	0.1	0.2	0.3	1.7	0.8	0.6	0.2	0.7	0.9	0.2
B′18	1.2	0.6	2.6	2.7	1.3	0.5	0.6	0.4	1.3	1.0	0.1
B′19	0.0	0.0	0.0	0.5	0.1	0.1	0.0	0.2	0.0	0.1	1.1
B′20	0.0	0.0	0.0	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.1
B′21	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.3
B′22	0.1	0.0	0.0	0.6	0.1	0.1	0.2	0.0	0.0	0.0	0.0
B′23	0.2	0.2	0.4	0.3	0.0	0.0	0.0	0.0	0.1	0.2	0.3
B′24	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.1	0.1	0.0

Table B10. PE population percentages along the vertical gradient at ~65°C average midday temperature site (EZ2; Appendix B Figure B3 D).

	% PE along 80 µm depth intervals Depth interval											
PE	**0/80µm**	**80/160µm**	**160/240µm**	**240/320µm**	**320/400µm**	**400/480 µm**	**480/560 µm**	**560/640 µm**	**640/720 µm**	**720/800 µm**	**800/880 µm**	**880/960 µm**
A1	12.4	13.5	18.3	17.7	4.7	19.2	21.5	19.3	17.9	10.9	7.1	8.0
A2	6.0	0.2	0.4	0.0	0.3	0.1	0.1	0.2	0.2	0.0	0.1	0.3
A3	0.2	1.5	1.0	1.0	0.0	0.3	0.1	0.0	0.1	0.0	0.0	0.1
A4	1.1	1.5	3.7	2.6	2.2	5.9	8.0	7.1	7.5	6.1	4.5	3.2
A5	3.8	1.2	2.9	4.4	1.1	3.0	3.2	4.0	1.4	1.7	0.7	1.7
A6	6.3	1.0	0.6	2.3	1.4	8.7	9.0	14.3	19.7	30.5	37.3	42.4
A7	4.9	0.1	0.5	0.5	0.8	0.2	0.2	0.1	0.0	0.0	0.0	0.5
A′8	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A′9	0.4	0.0	0.2	0.2	0.3	0.0	0.1	0.0	0.2	0.4	0.2	0.2
A10	2.5	1.5	0.5	1.0	0.0	0.2	0.2	0.0	0.3	0.1	0.1	0.0
A11	0.2	0.1	0.0	0.1	0.1	1.1	0.6	1.1	1.1	1.5	1.0	0.7
A12	3.3	1.1	2.4	3.7	0.5	3.4	3.5	2.2	2.6	0.9	0.6	0.4
A13	0.0	0.0	0.0	0.1	0.1	1.1	0.4	0.2	0.6	0.2	0.4	0.1
A14	6.9	2.0	0.9	2.5	1.2	11.2	12.6	19.2	21.9	35.6	39.4	31.3
A15	0.2	0.1	0.0	0.1	0.1	0.3	0.3	0.1	0.8	0.6	1.0	0.7
A′16	0.2	0.0	0.3	0.1	0.0	0.1	0.0	0.0	0.1	0.3	0.1	0.2
A′17	0.0	0.0	0.0	0.1	0.0	0.4	0.0	0.7	0.0	0.2	0.3	0.2
A′18	0.6	0.0	0.0	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.2
A19	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.2	0.0
A20	0.2	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2
A′21	0.0	0.0	0.3	0.8	0.0	0.2	0.3	0.2	0.2	0.9	0.9	0.5
A′22	0.0	2.3	0.0	0.5	0.0	0.5	0.0	0.0	0.0	0.0	0.4	0.0
B′1	0.0	0.4	0.3	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
B′2	3.8	8.2	3.1	3.8	7.5	3.7	3.1	3.6	2.7	1.7	0.6	1.7
B′3	1.9	1.5	3.1	0.5	1.6	0.3	0.2	0.1	0.2	0.1	0.0	0.0
B′4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′5	0.0	0.0	0.1	0.1	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
B′6	0.5	0.1	0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′7	1.2	1.0	0.3	0.1	2.6	0.0	0.0	0.0	0.1	0.0	0.0	0.0
B′8	4.7	3.6	1.6	1.6	5.8	2.6	2.1	1.3	1.7	1.4	0.5	0.9
B′9	13.5	20.3	26.7	21.6	26.1	11.3	9.9	5.2	5.3	1.4	1.4	2.0
B′10	0.0	0.2	0.0	0.3	0.3	0.2	0.0	0.6	0.0	0.0	0.0	0.2
B′11	0.7	0.6	0.5	0.3	0.5	0.1	0.2	0.0	0.1	0.0	0.1	0.0

B′12	3.6	9.1	7.2	4.8	4.4	1.6	1.4	1.5	1.7	1.1	0.5	0.5
B′13	0.0	0.4	0.0	0.0	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′14	0.0	0.2	0.0	0.1	0.7	0.0	0.2	0.0	0.0	0.0	0.0	0.0
B′15	4.7	2.0	4.3	3.4	0.5	1.7	3.8	0.6	0.8	0.1	0.3	0.3
B′16	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′17	0.0	0.5	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′18	0.5	1.0	0.8	0.7	0.5	0.2	0.2	0.4	0.0	0.1	0.1	0.0
B′19	0.0	0.8	1.0	1.0	0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0
B′20	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′22	0.7	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.1
B′23	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
B′24	0.0	0.2	0.1	0.0	0.2	0.1	0.0	0.2	0.2	0.0	0.1	0.0

Table B11. PE population percentages along the vertical gradient at the ~65°C average midday temperature site (EZ2; Figure 3C).

	% PE along 80 µm depth intervals Depth interval								
	0/80 µm	**80/160 µm**	**160/240 µm**	**240/320 µm**	**320/400 µm**	**400/480 µm**	**480/560 µm**	**560/640 µm**	**640/720 µm**
A1	35.4	35.1	33.8	26.0	27.3	22.4	30.7	26.0	25.2
A2	0.4	0.5	1.2	1.3	0.6	0.7	1.0	0.9	0.7
A3	0.9	1.4	0.8	0.4	0.2	0.2	0.2	0.1	0.3
A4	0.5	2.9	2.2	3.5	3.8	2.9	3.7	2.6	2.6
A5	7.1	7.2	3.7	6.3	3.1	2.5	1.7	3.5	12.6
A6	4.2	7.6	9.6	15.3	20.9	20.6	14.1	10.8	8.0
A7	14.4	12.9	7.8	8.8	10.7	8.5	11.2	12.9	11.7
A′8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A′9	0.0	0.0	0.4	0.3	0.2	0.2	0.6	1.7	0.0
A10	1.2	1.3	0.3	0.1	0.1	0.2	0.1	0.2	0.1
A11	0.1	0.1	0.1	0.6	0.3	1.1	0.2	0.5	0.5
A12	24.2	19.5	24.3	22.4	19.0	23.7	22.1	28.7	29.7
A13	0.4	0.1	0.2	0.1	0.2	0.2	0.0	0.1	0.1
A14	2.6	4.0	4.2	7.5	8.1	7.3	6.3	4.7	4.3
A15	0.4	0.4	0.5	0.6	0.2	0.4	0.4	0.1	0.7
A′16	0.2	0.1	0.2	0.3	0.2	0.4	0.5	0.6	0.3
A′17	0.2	0.2	0.1	0.1	0.2	0.6	0.7	1.3	0.0
A′18	2.5	2.6	0.9	1.6	1.3	1.0	0.7	2.1	3.2
A19	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0
A20	0.0	0.1	0.2	0.1	0.0	0.0	0.3	0.2	0.3
A′21	0.2	0.0	0.4	0.7	0.2	1.0	0.9	3.3	0.5
A′22	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′2	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
B′3	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0
B′4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′5	0.0	0.4	1.5	2.2	1.3	1.3	0.8	0.6	0.1
B′6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′7	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.1	0.1

B′8	0.7	0.5	0.4	0.8	0.2	0.7	0.8	0.2	0.7
B′9	2.9	2.9	3.3	1.7	1.7	2.9	2.3	1.3	1.4
B′10	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0
B′11	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0
B′12	0.9	1.0	0.7	0.0	0.0	0.3	0.1	0.3	0.3
B′13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′14	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.1	0.0
B′15	2.3	2.3	2.6	0.6	0.9	0.7	0.7	0.4	2.0
B′16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′18	0.1	0.2	0.2	0.2	0.1	0.0	0.1	0.1	0.0
B′19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′22	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′23	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′24	0.1	0.3	0.0	0.2	0.0	0.0	0.1	0.2	0.0

Table B12. Abundant PE population transcript abundance percentages averaged between replicate samples (68°C was not replicated) during the morning and evening light transition periods in EZ1, EZ2 and EZ3. PEs not listed were not detected in the analysis.

	Time												
PE	**% PE at 63°C (EZ1)**				**% PE at 65°C (EZ2)**					**% PE at 68°C (EZ3)**			
	0700	**0800**	**0930**	**1100**	**0700**	**0800**	**0930**	**1300**	**1830**	**0800**	**1100**	**1300**	**1830**
A1	8.8	37.8	9.9	11.5	40.1	42.6	41.0	48.0	30.3	0.0	0.0	0.0	0.0
A2	0.0	0.9	3.3	0.0	0.3	0.0	5.7	7.9	0.0	0.0	0.0	0.0	0.0
A4	1.2	4.5	3.0	7.5	0.8	0.3	3.9	1.3	5.0	0.0	0.0	0.0	0.0
A6	0.9	10.5	8.2	2.0	3.8	13.2	7.6	5.3	13.6	0.0	0.0	0.0	0.0
A7	0.0	2.0	1.1	0.2	8.3	3.7	6.6	7.7	0.0	0.0	0.0	0.0	0.0
A′8	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	31.1	67.3	43.3	14.3
A′9	2.6	0.0	0.0	0.0	3.0	2.9	8.4	3.8	0.5	23.4	21.8	17.8	42.9
A15	0.0	6.1	1.4	0.2	14.3	16.9	7.8	14.4	4.1	0.0	0.0	0.0	0.0
A′16	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.9	13.6	10.4	10.6	42.9
A′17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	31.9	0.5	28.3	0.0
B′2	14.1	6.4	9.1	16.8	1.2	2.3	1.7	0.0	3.8	0.0	0.0	0.0	0.0
B′6	0.9	3.7	1.9	0.4	15.6	7.2	2.0	0.0	22.6	0.0	0.0	0.0	0.0
B′7	3.0	2.4	8.4	5.1	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′8	24.9	6.8	11.9	13.1	2.7	1.3	4.1	1.9	5.9	0.0	0.0	0.0	0.0
B′9	24.1	8.4	21.2	21.9	2.8	1.5	5.1	3.0	3.6	0.0	0.0	0.0	0.0
B′11	3.9	0.0	2.2	2.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table B13. PE population percentages changing temporally during the temperature shift from EZ1 to EZ2. Samples were collected between 1200 and 1300 h.

	% PE over time					
	Day 0	**Day 2**	**Day 4**	**Day 6**	**Day 2**	**Day 4**
A1	4.3	24.5	36.0	16.8	23.4	18.2
A2	0.0	0.0	0.1	0.8	0.3	0.0
A3	0.1	0.1	0.9	0.3	0.5	0.1
A4	5.3	3.6	13.4	4.7	2.7	7.2
A5	3.0	3.3	20.5	4.1	4.4	3.2
A6	13.2	7.9	8.9	12.2	5.0	6.9
A7	0.0	0.3	3.4	0.3	1.0	0.3
A′8	0.0	0.0	0.0	0.1	0.0	0.0
A′9	0.0	0.3	0.0	0.0	0.3	0.3
A10	0.1	0.4	0.1	0.2	0.4	0.1
A11	0.2	0.0	1.0	0.3	0.0	0.8
A12	0.1	0.6	1.0	0.5	0.5	1.7
A13	0.0	0.0	0.0	0.2	1.0	0.2
A14	3.8	15.6	5.1	14.6	4.9	15.8
A15	0.1	0.2	0.4	0.2	0.2	1.0
A′16	0.0	0.2	0.0	0.2	0.0	0.0
A′17	0.0	0.0	0.0	0.0	0.0	0.0
A′18	0.0	0.0	1.3	0.1	0.1	0.1
A19	0.0	0.0	0.0	0.0	0.0	0.0
A20	0.0	0.0	0.0	0.2	0.0	0.0
A′21	0.0	0.0	0.3	0.0	0.2	0.0
A′22	0.0	0.0	0.1	0.0	0.3	0.2
B′1	0.4	0.0	0.0	0.0	0.2	0.0
B′2	10.2	1.7	0.0	2.2	4.1	2.9
B′3	0.2	0.2	0.2	0.3	0.2	0.3
B′4	0.0	0.0	0.0	0.0	0.1	0.0
B′5	0.0	0.9	0.3	0.6	1.4	1.3
B′6	0.0	0.0	0.0	0.0	0.1	0.0
B′7	0.4	0.0	0.6	0.2	0.5	0.1
B′8	8.7	7.4	1.2	7.2	4.1	4.7
B′9	23.7	9.7	3.6	17.5	13.3	13.4
B′10	0.2	0.2	0.0	0.2	0.3	0.0
B′11	0.5	0.5	0.0	0.4	0.2	0.4
B′12	3.8	0.4	0.3	1.2	2.8	2.2
B′13	0.1	0.0	0.0	0.0	0.0	0.1
B′14	0.0	0.0	0.1	0.1	0.5	0.0
B′15	0.4	8.2	1.1	2.8	12.6	6.3
B′16	0.0	0.0	0.0	0.0	0.0	0.0
B′17	0.5	0.6	0.0	0.1	0.3	0.1
B′18	0.7	0.1	0.0	0.3	0.5	0.9
B′19	0.1	0.0	0.0	0.0	0.0	0.0
B′20	0.0	0.3	0.6	0.0	0.0	0.0
B′21	0.0	0.0	0.0	0.0	0.0	0.0
B′22	0.0	0.2	0.0	0.1	0.1	0.1
B′23	0.1	0.0	0.0	0.1	0.0	0.0
B′24	0.1	0.1	0.2	0.2	0.3	0.1

Table B14. PE population percentages over time during the light alteration experiments conducted in EZ2. Samples were collected between 1200 and 1300 h.

	Time						
	Light reduction					UV removal	
	Replicate sample 1			Replicate sample 2			
PE	**Day 0**	**Day 2**	**Day 4**	**Day 2**	**Day 4**	**Day 2**	**Day 4**
A1	13.0	20.6	19.5	14.5	13.4	10.4	10.0
A2	8.4	5.5	3.8	5.8	5.2	3.6	4.1
A3	0.3	0.1	0.1	0.1	0.3	0.0	0.0
A4	8.2	3.2	2.5	1.7	0.3	2.2	1.2
A5	5.4	6.1	4.6	4.8	8.7	1.5	3.2
A6	24.2	11.6	9.3	13.5	11.5	7.9	11.5
A7	7.3	18.1	10.1	7.6	7.5	4.6	7.1
A′8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A′9	0.0	0.2	0.7	0.3	0.8	0.1	0.3
A10	1.2	1.6	0.3	4.0	1.0	0.3	1.0
A11	2.5	1.3	2.4	0.5	0.8	1.9	0.3
A12	5.9	16.2	10.2	9.1	20.3	6.3	6.3
A13	0.1	0.2	0.1	0.0	0.0	0.2	0.2
A14	5.2	1.8	1.8	3.5	7.6	2.1	0.9
A15	0.1	0.0	0.1	0.2	0.5	0.0	0.0
A′16	0.0	0.0	0.2	0.0	0.0	0.0	0.0
A′17	0.0	0.1	0.1	0.1	0.1	0.0	0.0
A′18	2.4	5.3	2.8	2.5	3.2	0.6	1.6
A19	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A20	0.0	0.0	0.1	0.0	0.0	0.0	0.0
A′21	0.0	0.1	0.4	0.0	0.3	0.1	0.0
A′22	2.0	1.4	0.7	0.7	0.2	0.3	0.6
B′1	1.7	0.0	1.1	0.8	0.1	1.1	2.0
B′2	0.0	0.0	0.1	0.0	0.1	0.1	0.3
B′3	0.0	0.0	0.0	0.0	0.0	0.1	0.0
B′4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′5	0.2	0.0	0.0	0.3	0.3	0.4	0.2
B′6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′7	0.2	0.1	0.4	0.4	0.2	0.6	0.4

B′8	0.4	0.0	0.5	0.4	1.0	0.6	0.4
B′9	4.5	2.0	5.1	2.9	3.1	17.1	14.8
B′10	0.1	0.0	0.0	0.4	0.2	0.6	0.4
B′11	0.1	0.0	0.3	0.3	0.2	0.4	0.0
B′12	0.8	0.1	1.2	1.0	0.0	1.1	1.3
B′13	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′14	1.5	1.7	5.7	3.1	2.6	5.9	6.1
B′15	5.6	1.9	10.4	17.1	4.0	16.2	14.3
B′16	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′17	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′18	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′19	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′20	0.1	0.7	1.0	3.6	0.6	0.3	0.1
B′21	0.0	0.0	0.4	0.0	0.0	0.0	0.0
B′22	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′23	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′24	0.4	0.0	0.1	0.0	0.0	0.4	0.2

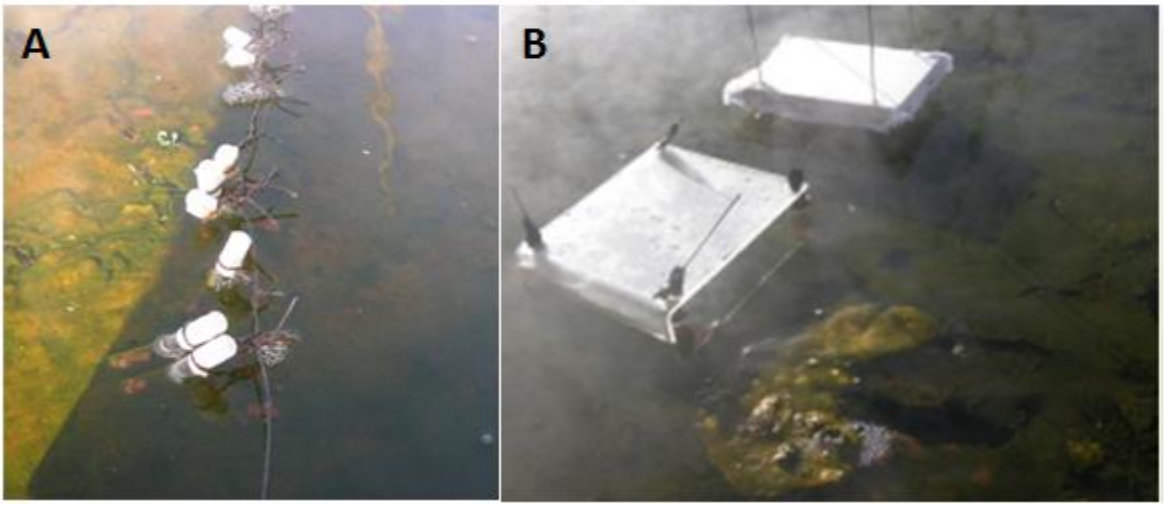


Figure B1. Photographs of the temperature shift experiment and light alteration covers. (A) 3 ml vials suspended by an aluminum wire, used to suspend samples in spring water for temperature-shift experiments. (B) Muslin light-reducing cover and Plexiglas UV-blocking cover used in light alteration experiments. Both experiments were conducted in EZ2.

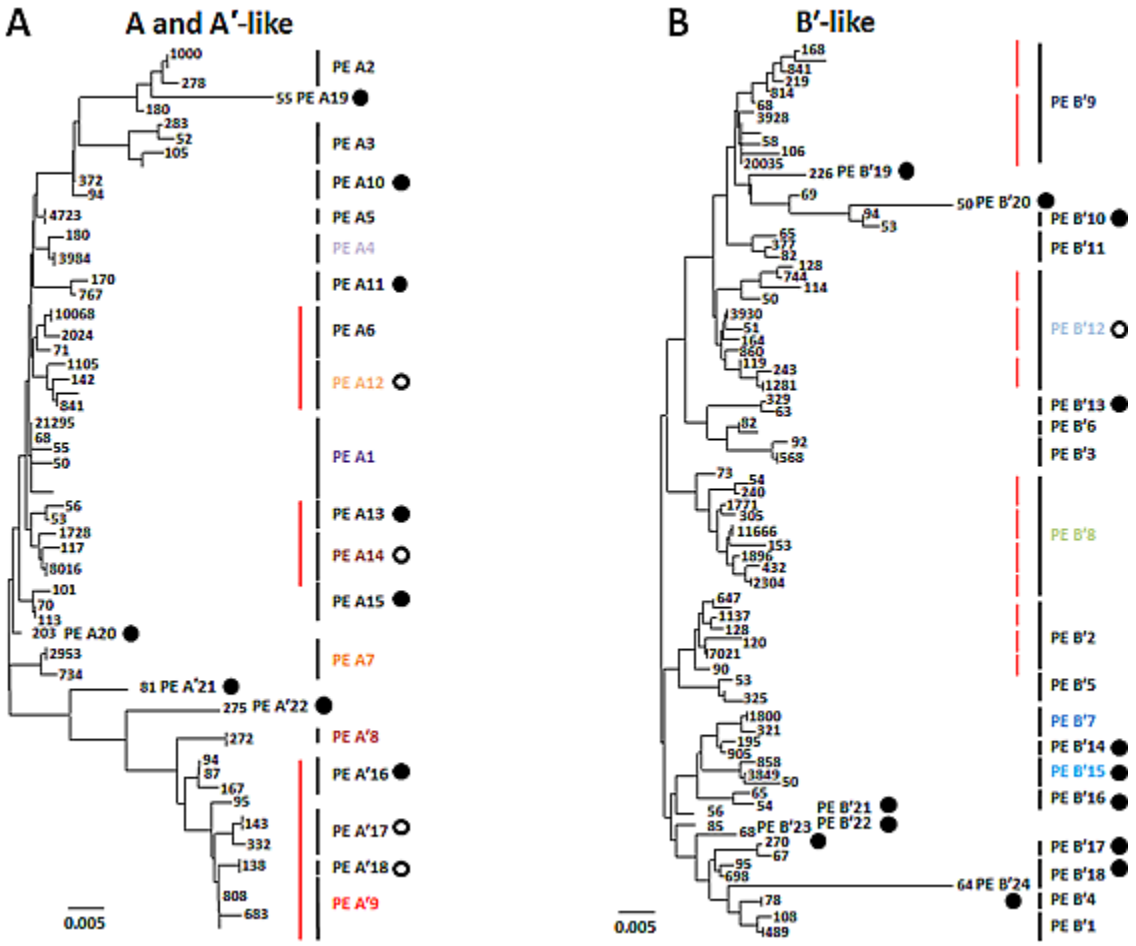


Figure B2. Comparison of putative ecotypes (PEs) demarcated from high-frequency *psaA* sequences by Ecotype Simulation using the fine-scale or conservative demarcation approaches. (A) A/A′ lineage. Here, black vertical bars are PEs demarcated using the fine-scale approach (as in main text Figure 1A) and red vertical bars are PEs demarcated using the conservative approach. (B) B′ lineage. Here, black vertical bars are PEs demarcated using the conservative approach (as in main-text Figure 1B) and red vertical bars indicate PEs demarcated using the fine-scale approach. Scale bars represent (A) 0.005 and (B) 0.005 substitutions per site.

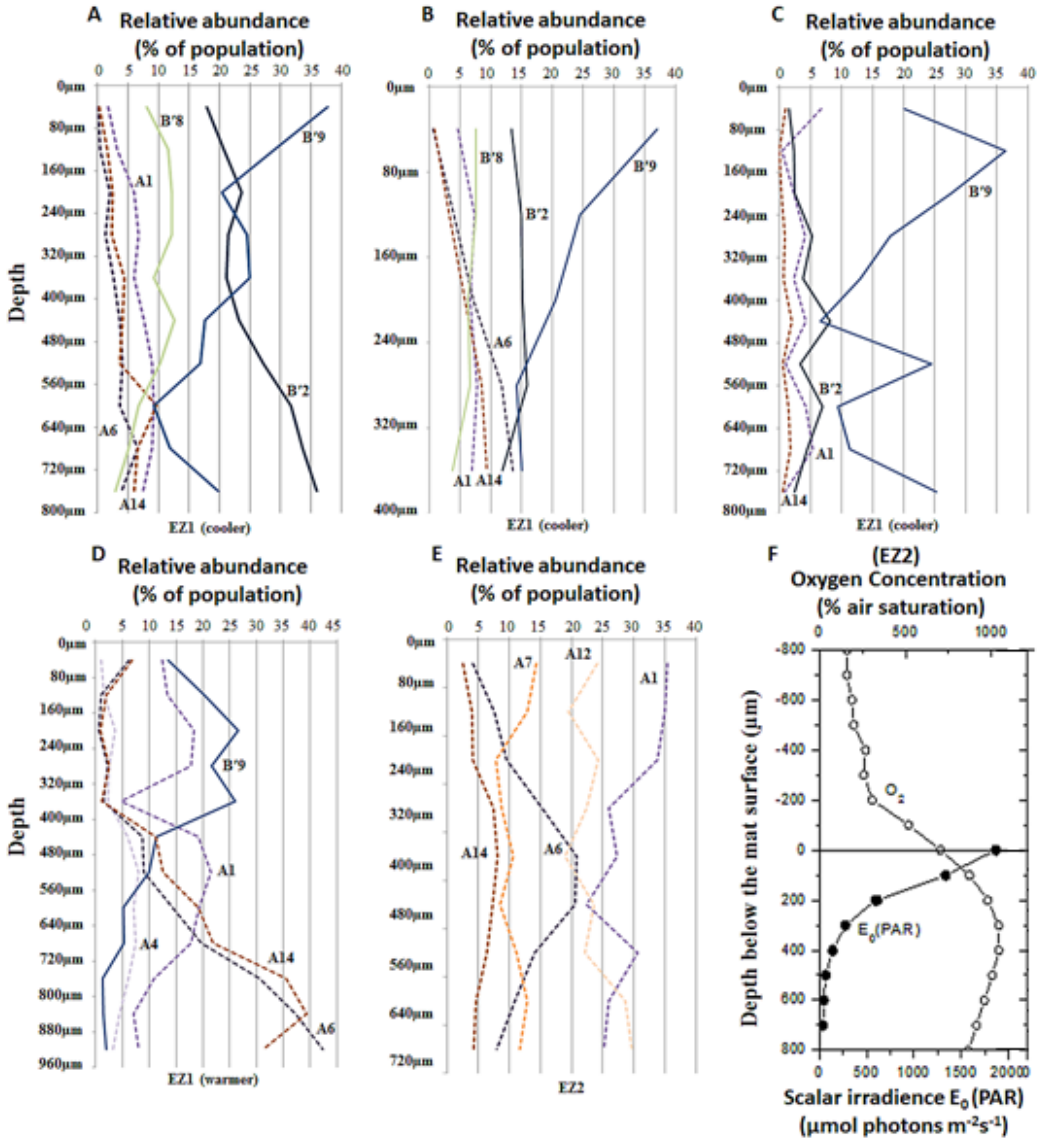


Figure B3. Relative abundance of abundant putative ecotypes (PEs) in ~80 µm-thick subsections along the vertical dimension of cores ostensibly collected at 60°C (A and B; also see Appendix B Tables B6 and B7), at 63°C (C; Appendix B Table B8) and at 65°C (EZ1) (D; Appendix B Table B10) and 65°C (EZ2) (Appendix B Table B11) relative to light (PAR) and oxygen profiles (F). Solid lines represent B′-like PEs and dashed lines represent A-like PEs. Ecological zone (EZ) assignments were made after determining the PE composition of pooled vertical intervals of individual cores, which were compared to whole-cores used in the temperature distribution analysis (see Appendix B Table B5).

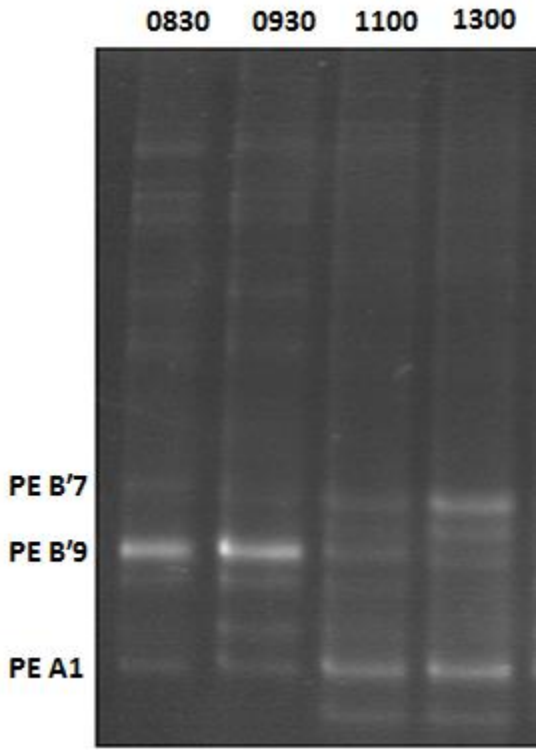


Figure B4. Denaturing gradient gel electrophoresis (DGGE) gel showing amplified *psaA* cDNA from September 2006 60°C samples collected over the morning light transition period, corresponding to transcript patterns in EZ1 from September 2008.

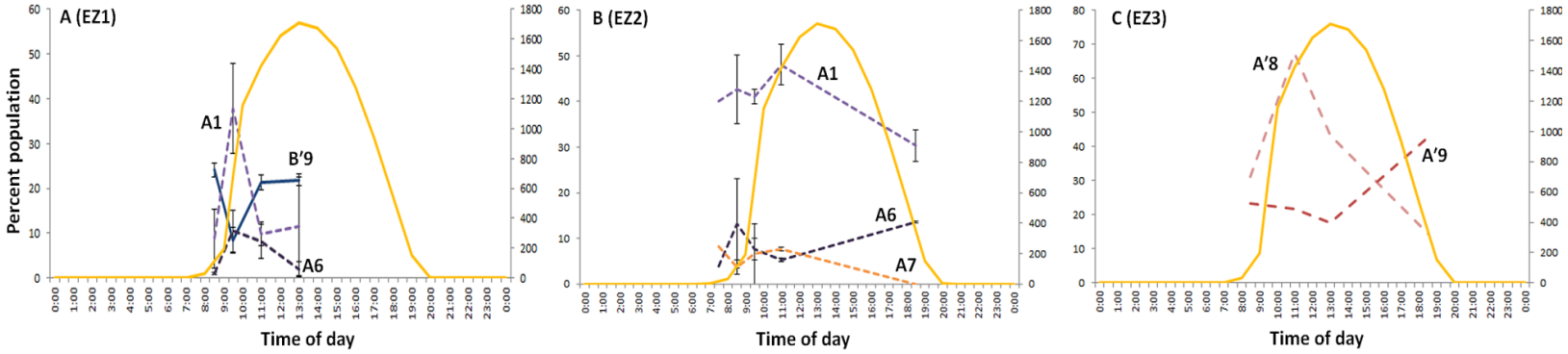
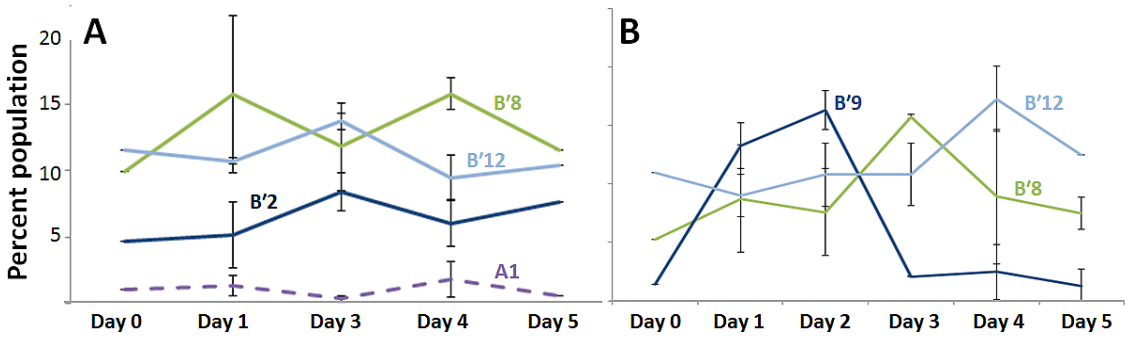


Figure B5. Downwelling photon irradiance (yellow line) and relative abundances of abundant putative ecotype (PE) transcripts measured over the morning and evening light transition periods on 13, 14 and 15 September 2008 in ecological zone (EZ) 1 (A) and EZ2 (B), and on 12 and 13 September 2009 in EZ3 (C). Solid lines indicate B′-like PEs, dashed lines indicate A-like PEs and longer dashed lines represent A′-like PEs. Samples from evening time points at 1830 h were collected on the previous day. Bars in A and B indicate the range between replicate samples.



Appendix Figure B6. Percent population of putative ecotypes in (A) control samples and (B) changing in relative abundance in response to UV exclusion over a five day period in ecological zone 1. Experiments were conducted in 1996 by David Ward. Bars indicate the range between replicate samples.

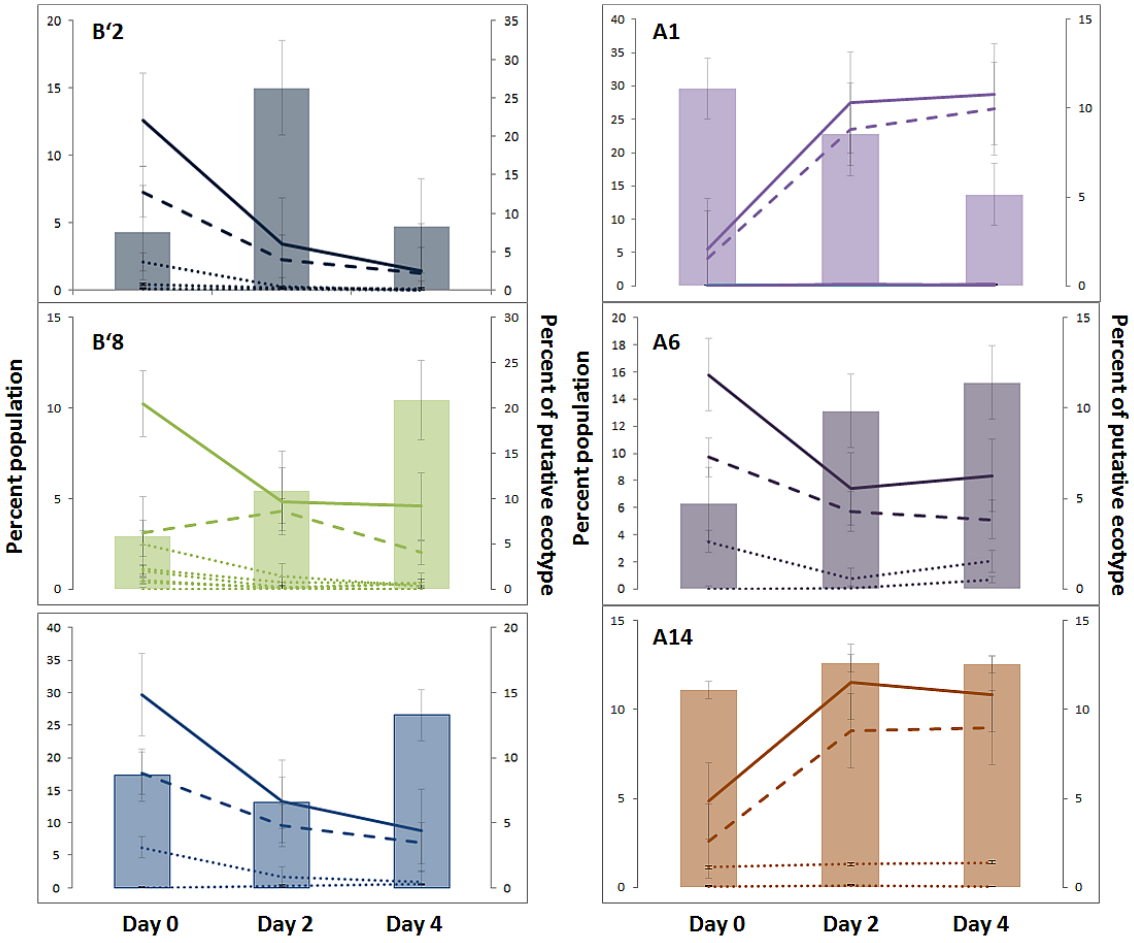


Figure B7 Change in relative abundance of abundant putative ecotypes (PEs; solid lines) over a four day period after shifting samples from ecological zone (EZ) 1 to EZ2 (left axis). Bars represent the percentage of low frequency sequences (≤50 total copies in all samples) associated with the total number of sequences in a PE (right axis). PE dominant variants are represented by dashed lines, and associated high-frequency sequences are represented by dotted lines. Bars represent standard error of replicate samples. Samples correspond to temperature shift samples in main-text Material and Methods and Figure 5A.

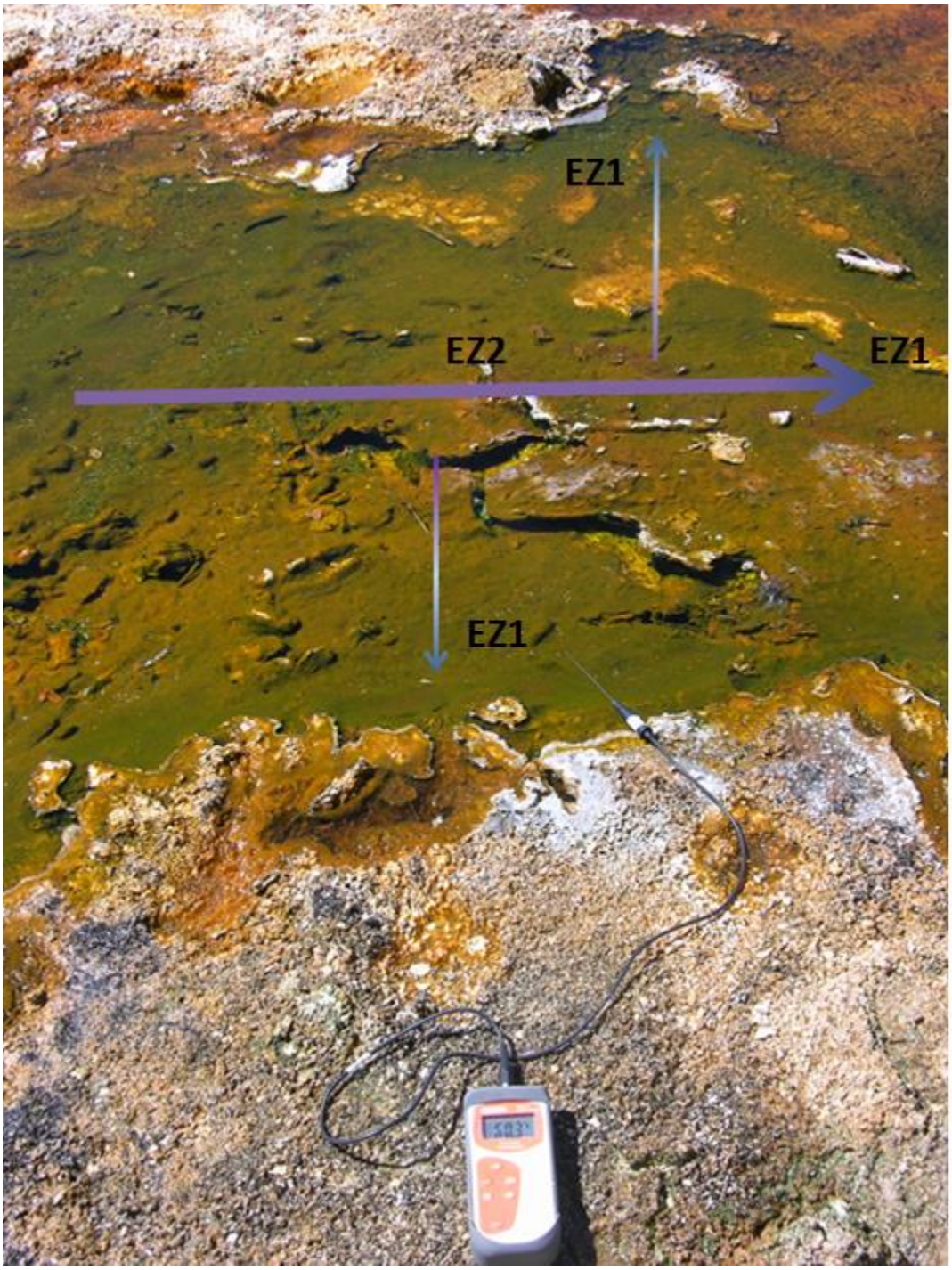


Figure B8. Temperature distribution allong the effluent flow path in Mushroom Spring, Yellowstone National Park. Large arrow indicates main flow path from ecological zone (EZ) 2 (left; purple) to EZ1 (right; blue). Smaller arrows indicate perpendicular temperature gradient from the main flow path.

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

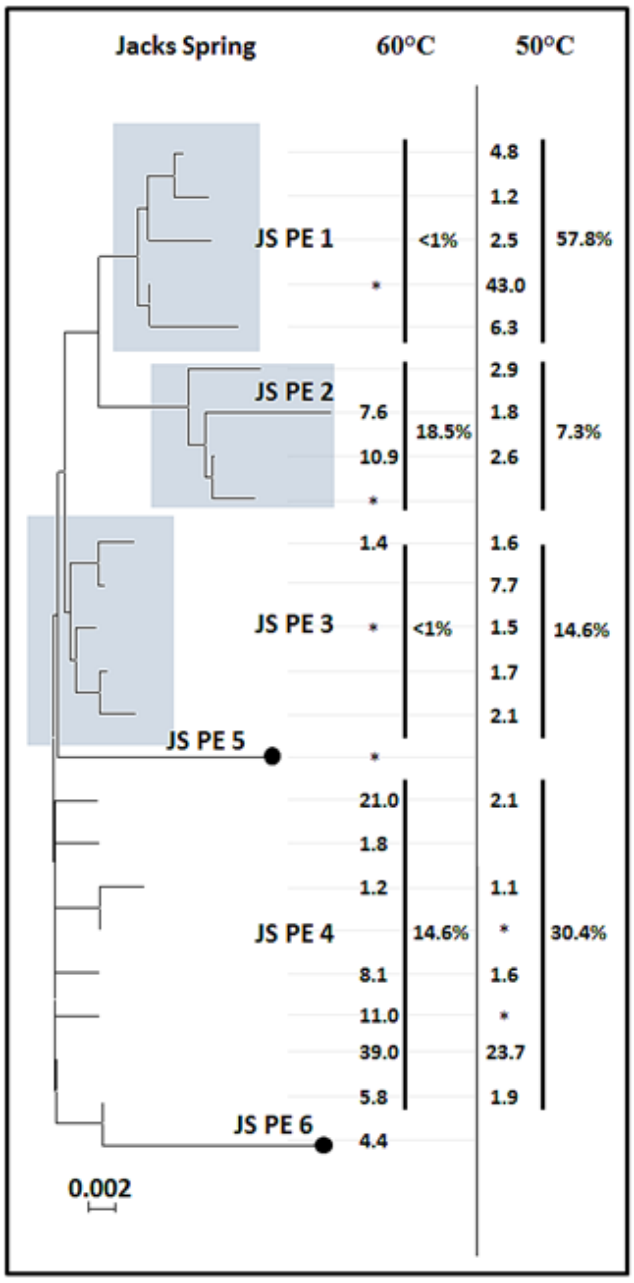


Figure C1. Phylogeny of high-frequency sequences (HFSs) ≥1% of the sample at 2 ecological settings at Jack's Spring (JS). JS samples corresponded to average temperatures of 60°C (left) and 50°C (right). JS contains only A-like HFSs. Vertical lines indicate putative ecotypes (PEs) demarcated by Ecotype Simulation; black dots represent singleton HFS PEs. Shaded boxes indicate mostly ecological site-specific HFS clades. Scale bar indicates 0.002 nucleotide substitutions per site.

APPENDIX D

SUPPLEMENTARY INFORMATION FOR CHAPTER 6

Section I: Ecological Population Responses to Perturbation in Ecological Zone 1

Introduction

As a compliment to the main-text disturbance experiments, the responses of *psaA*-defined species populations to altered light environments were investigated, effectively providing an analysis of perturbation responses similar to those reported in Chapter 3, although in a different ecological zone (EZ) (EZ1).

Materials and Methods

Experimental design, sampling, molecular methods and ecotype demarcations were as described as in the main text. Results for relatively abundant PEs (>5% of total PE abundance) are presented in Appendix D Figure D1, but all results are presented in Appendix D Tables D2-D7. Standard deviations are reported in Appendix D Table D11.

Results and Discussion

Studies of responses to environmental change under different light conditions allowed me to observe the population dynamics of putative ecotypes (PEs) in response to perturbations in EZ1, and to compare PE responses to those previously observed in EZ2 (Chapter 3, Figure 3.5).

~92% Ambient Light Reduction. Compared to unperturbed control samples (Appendix Figure D1 A) after photon irradiance was reduced by ~92%, B′-like PEs B′8, B′12 and ND PE B′5 increased in relative abundance between days 1 and 3, though ND

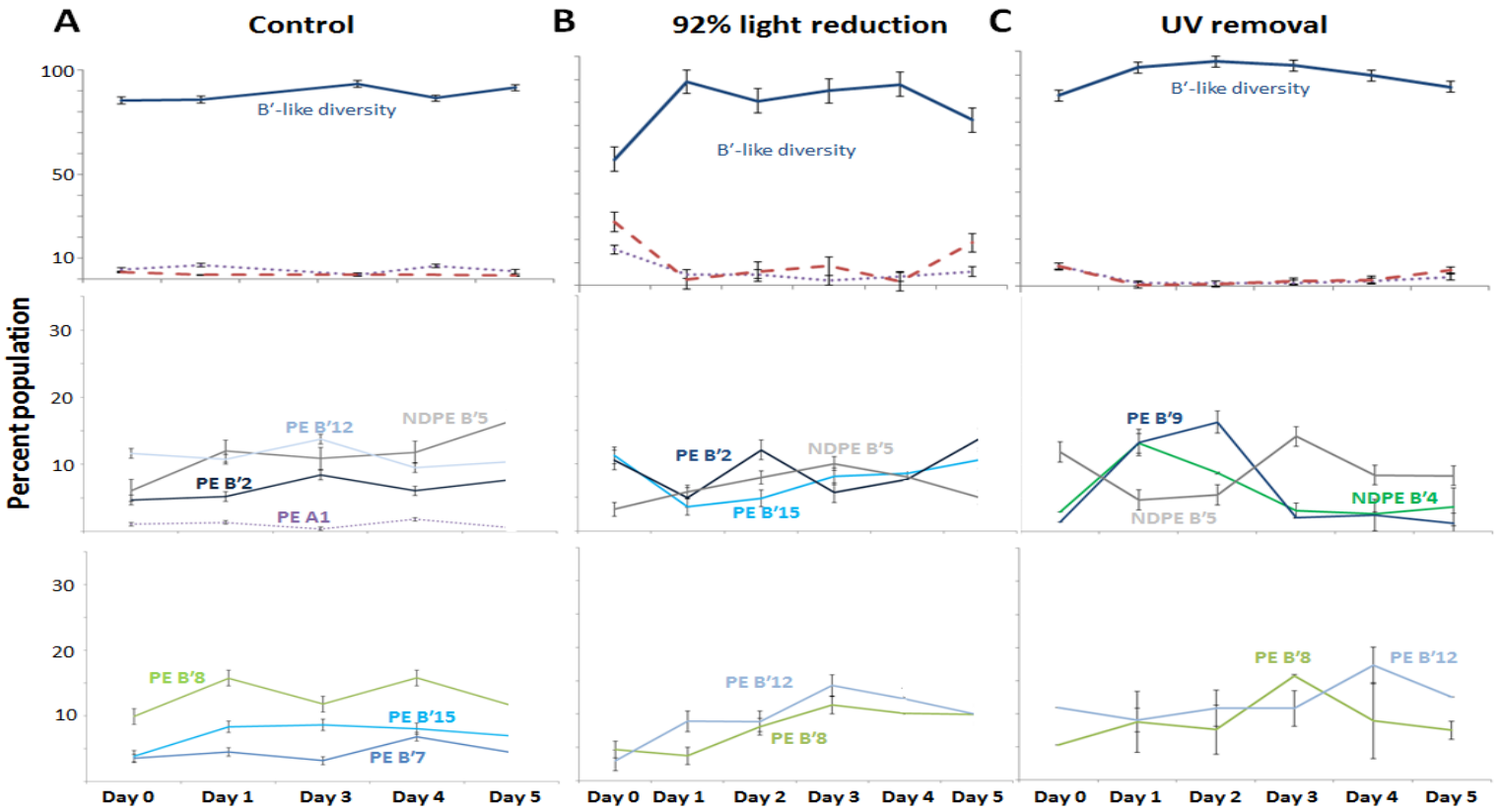


Figure D1. Population percentages of abundant *Synechococcus* putative ecotypes (PEs) in samples collected over a 5 day period in undisturbed Mushroom Spring in ecological zone (EZ) 1 (~58°C) under different light regimes; (A) Ambient light, (B) ~92% light reduction and (D) UV-light exclusion conditions in a section of mat at Mushroom Spring. Change in total A′, A and B′-like diversity (top); solid lines represent B′-like PEs, dashed lines represent A′-like PEs and dotted lines represent A-like PEs. Middle and bottom panels report different predominant PEs for the same experimental conditions.

PE B′5 decreased again between days 3 and 5 (Appendix Figure D1 B). PEs B′2 and B′15 immediately decreased in relative abundance after light reduction, though these PEs increased again in relative abundance between days 1-2 and 5. For reference, PEs B′8 and B′12 were both located at the mat surface in EZ1 (Chapter 3, Figure 3.3A), and PE B′2 was more relatively abundant in the middle and deeper mat layers, though it was also present in surface mat layers in lower relative abundance (Chapter 3; Appendix B Figure B3 A and B).

The light reduction experiment described in Chapter 3 was conducted in 2008 in EZ2, where A-like PEs are predominant. In that experiment, deep-subsurface populations PEs A4 and A6 declined in relative abundance, while the surface-oriented populations PEs A1 and A12 increased in relative abundance (Chapter 3; Figure 3.5B). Surface and subsurface B′-like PE populations in EZ1 (1996) responded similarly to A-like PE populations in EZ2 (2008) to light reduction (compare Figures 3.5B and 6.6A). The increase of surface-oriented PEs, and the immediate decline in relative abundance of PE B′2 in response to light reduction, was similar to A-like populations in EZ2 (Chapter 3, Figure 3.5B). The subsequent increase could be a result of PE B′2 inhabiting depth intervals closer to the mat surface in EZ1 than PEs A4 and A6 in EZ2 (Chapter 3, Appendix B Figure B3).

UV-light Exclusion. As in the light reduction experiment, B′-like sequence diversity was predominant in all UV-blocked samples throughout the duration of the study. After the exclusion of UV-light, ND PE B′5 immediately decreased in relative abundance, though it increased again at day 3 (Appendix Figure D1 C). PEs B′9 and ND

PE B′4 immediately increased in relative abundance, despite being low-abundance PEs initially (i.e., <1% contribution to the *psaA* pool) in the mat at this site. After 3 days PEs B′8 and ND PE B′5 had also increased in relative abundance, and PEs B′9 and ND PE B′4 sharply decreased in relative abundance. At days 4 and 5, PE B′12 increased to the most relatively abundant population in those samples.

In 2008 studies (Chapter 3; Figure 3.5C), surface-oriented PEs B′9 and B′15 increased in relative abundance after 2 days when UV-light was removed from EZ2, despite being low relative abundance populations initially, though after 6 days both PEs had declined in relative abundance to near zero (data not shown). PEs responded similarly to UV-light exclusion in EZ1 in 1996 (compare Figures 3.5C and 6.6A). This could be because these B′-like PE populations are better adapted to minimal UV environments, or spend less energy protecting themselves from the harmful effects of UV, and have an initial growth advantage when UV is excluded. However, after a period of time other PE populations that were originally displaced increased in relative abundance, possibly indicating an end to the initial and temporary advantage gained after the exclusion of UV-light. Also, condensation on the light altering screens may have caused light scattering, which could have influenced changes in PE relative abundances during the sampling period. Mean percent population, standard deviation and responses to environmental perturbations of all predominant PEs (≥5% in a sample) are compared in Appendix D Table D1. Clear differences in PE responses under the different light conditions are evident.

Table D1. Summary of abundant putative ecotypes (PEs) ≥5% of a sample after environmental perturbation. Known surface-oriented PEs are indicated in green. The mean percent population and percent standard deviation (%STD; in parentheses) are reported; arrows indicate an increase in PE relative abundance in response to perturbation at a time point during the duration of the study.

PE	Control	~92% µE [a]	-UV
B′2	6.4(45)	9.1(56)	
B′3			
NDB′4			5.7(16)↑
NDB′5	4.5(18)	6.7(72)	8.8(13)
B′7	4.5(57)		
B′8	13(34)	8(42)↑	9(22)
B′9			6.1(28)↑
B′12	11(12)	9.6(34)↑	12(11)
B′15	7.2(49)	7.8(29)	
A1	1(89)		
%STD	35-53%	19-84	14-30

Conclusions

PE responses to environmental perturbations in EZ1 were similar to those observed in EZ2 in previous studies (Chapter 3, Figure 3.5). Surface-oriented B′-like PEs immediately increased in relative abundance in response to UV-light removal, which was similar to previous UV-light exclusion studies (Chapter 3; Figure 3.5C). Surface-oriented populations increased in relative abundance in response to light reduction, and subsurface populations declined in relative abundance, possibly because light was reduced below a threshold intensity required for growth.

Section II: Temperature Effects on PE Populations Transferred Upstream From EZ1 to EZ3

Introduction

The disturbance experiments conducted in 1996 by David Ward also included a temperature-shift experiment of Mushroom Spring mat samples from ecological zone (EZ) 1 to EZ3. In 1996, the only insight into the effect of temperature shifts on A/B-like *Synechococcus* populations was from an oligonucleotide probe study of the effects of temperature on 16S rRNA genotypes (Ruff-Roberts et al., 1994), which was conducted before A′-like *Synechococcus* populations had been discovered by Ferris and Ward (1997). In order to increase the molecular resolution of previous studies the 1996 temperature-shift samples were analyzed using *psaA* variation, as reported here.

Materials and Methods

To analyze the effects of temperature on mat *Synechococcus* ecological populations, Mushroom Spring mat samples were transferred from a downstream site at ~58°C (55.9 to 59.4°C), to an upstream site at ~67°C (66.1 to 67.3°C). Samples were collected periodically between 21 June and 10 July 1996 with a #4 cork borer (38.5 mm^2), and were placed in 4 dram vials filled with water from the site of collection, submerged in spring water in the source pool, and incubated for the duration of the experiment. Mat samples were also collected at the site of incubation (source pool) over the 15 day period, and all samples were immediately frozen on dry ice (-20°C) and stored at -80°C until analysis in 2011. All molecular methods and sequence analyses were as

described in the main text (Chapter 6, Materials and Methods; Appendix D Tables D8 and D9).

Results and Discussion

Control samples collected at the site of incubation in the source pool consisted of A′-like PE populations (Appendix Figure D2 C). However, A′-like PE diversity varied widely among samples over the 15 day period studied, possibly because of the source pool and/or the cellular content of the water contributing to the sample.

Before the downstream samples were transferred upstream (Day 0 in Appendix Figure D2 A and B), B′-like PEs dominated the mat sample, with PE B′8 being the most relatively abundant population (Appendix Figure D1 A). B′-like PEs were present over the first 4 days, though minimal fluctuations in PE relative abundance were observed. After 15 days ND PEs A′4 and A′5 and PE A′9 and increased in relative abundance, and B′-like PE diversity was no longer detected (Appendix Figure D2 A and B). Additionally, all A′-like PEs that had increased in relative abundance on day 15 consisted of only 1 HFS sequence without variation, indicating that these populations were clonal. The clonality of the growing A′-like populations could have been the result of these populations being in low-abundance in the vials. Despite limited diffusion within vials due to restricted flow, PE populations shifted in relative abundance to match previous flow path distribution studies (Chapter 3; Figure 3.2), suggesting that temperature is a strong selective force in this mat system.

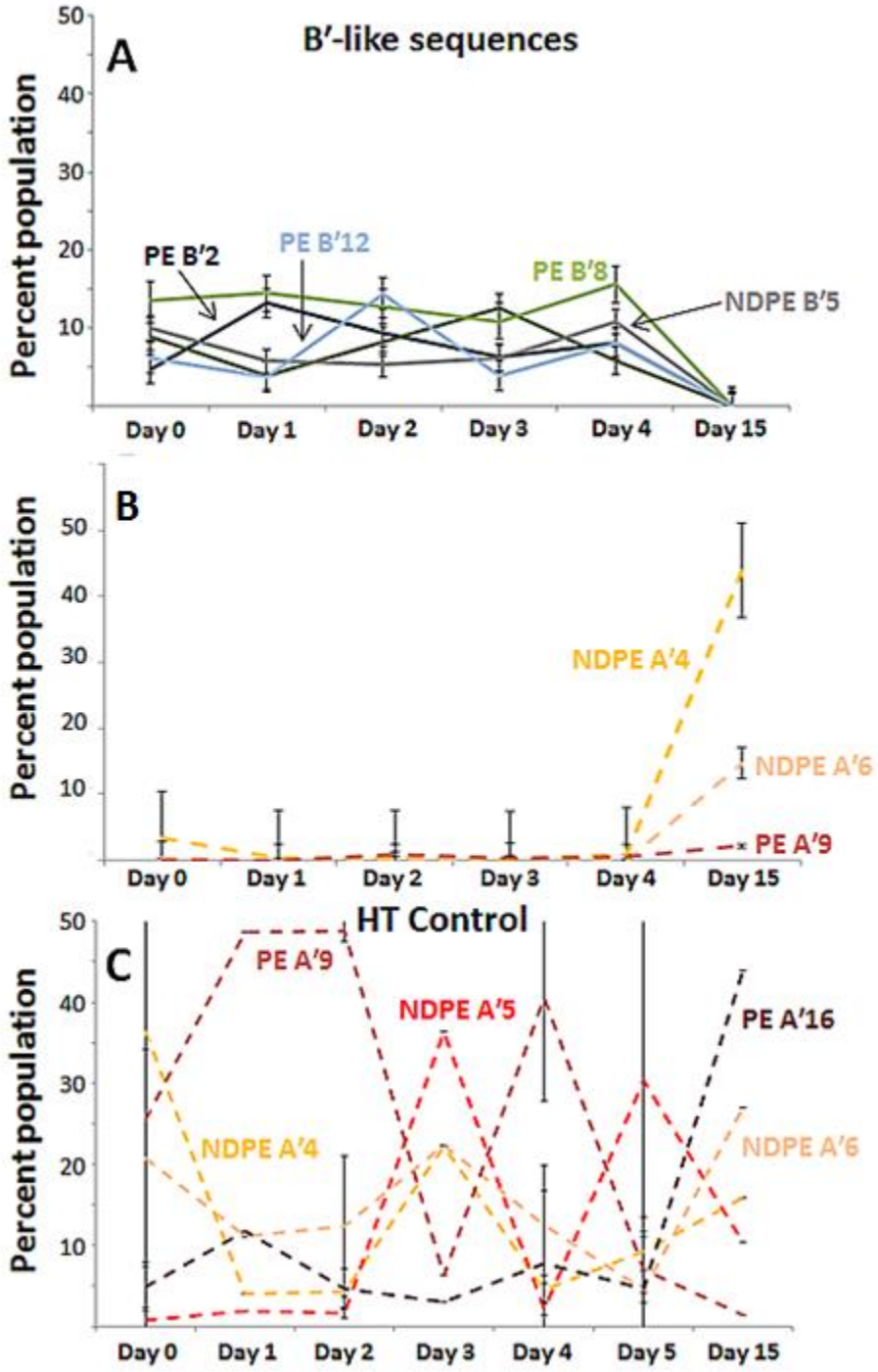


Figure D2. Relative abundances of abundant putative ecotypes (PEs) over a 15 day period of (A) B′-like PEs and (B) A′-like PEs after shifting samples collected from ecological zone (EZ) 1 (~58°C) to EZ3 (~67°C) in Mushroom Spring. (C) Unaltered mat samples collected at a ~67°C site (EZ3). Error bars represent standard error between replicate samples. Missing replicates at some time points are due to failed PCR or sequencing reactions.

In Ruff-Roberts et al. (1994), after moving samples from a ~50°C site (EZ1) upstream to a ~65°C site (EZ2), RNA, rather than DNA, was extracted and analyzed. 16S rRNAs with B′-like genotypes were no longer present after 7 days, and were replaced by A-like 16S rRNA genotypes that are typically found in abundance in EZ2. However, the 16S rRNA-specific probes used by Ruff-Roberts et al. (1994) might not have distinguished between A and A′-like populations. In this experiment, *psaA* gene segments, as opposed to rRNA transcripts, were analyzed. B′-like PEs were no longer detected after day 4, and were replaced with A′-like PEs that are typically found in abundance in EZ3 by day 15, which was similar to what was observed by Ruff-Roberts et al. (1994), despite the experiments having been located in different EZs. The presence of B′-like PEs over the first 4 days in this experiment could have been due to the slower degradation rate of DNA compared to RNA, and/or to sequences from DNA in low-abundance dead or dying cells having amplified in PCR reactions before A′-like PEs could increase in relative abundance, possibly because of a several-day lag-phase before growth in the limiting conditions of the vial.

Conclusions

B′-like PE populations in samples transferred from EZ1 to EZ3 were replaced by A′-like populations typically found in EZ3 between 4 and 15 days, similar to previous studies by Ruff-Roberts et al. (1994), indicating that temperature is a powerful selective force determining population structure in MS.

Table D2. Population percentage of putative ecotypes (PEs) over a 5 day period (0-5) in control (C) (unaltered mat) samples. A and B indicate replicate samples where applicable.

PE	0A	1A	1B	3A	3B	4A	4B	5A
B'1	8.8	1.5	4.0	1.0	1.5	0.5	1.9	1.7
B'2	4.7	7.7	2.7	7.0	9.9	7.8	4.3	7.7
B'3	3.0	2.0	3.6	6.7	1.5	1.5	4.5	3.3
B'5	2.2	1.6	2.3	4.4	2.0	1.0	2.7	1.3
B'7	3.6	2.6	6.4	2.1	4.3	3.7	9.9	4.4
B'8	9.9	9.9	21.7	8.5	15.1	17.0	14.6	11.6
B'9	2.1	2.9	2.6	2.3	2.4	4.4	1.9	5.2
B'11	3.3	5.7	3.6	5.1	6.6	6.0	4.9	4.9
B'12	11.6	11.0	10.5	13.1	14.3	7.9	11.2	10.4
ND B'1	1.2	1.2	0.3	0.7	0.3	0.3	0.0	1.2
ND B'2	1.6	0.2	0.0	0.8	1.5	0.2	0.6	0.5
ND B'3	2.6	4.5	3.7	7.0	6.3	3.9	5.0	2.9
ND B'4	2.1	4.3	0.6	2.6	3.1	1.9	1.1	3.9
ND B'5	6.1	15.0	8.9	10.9	10.8	10.4	13.2	16.4
ND B'6	3.7	5.4	2.5	7.0	3.6	5.6	2.5	3.5
ND B'7	3.9	12.4	4.3	11.1	6.2	10.3	5.8	7.0
ND B'8	2.0	0.9	2.0	1.1	3.7	0.5	4.0	3.6
ND B'9	3.1	0.6	3.3	1.0	1.2	0.0	2.2	2.2
A1	1.1	0.6	2.1	0.1	0.6	3.1	0.5	0.6
A2	0.6	0.4	0.8	0.0	0.3	1.3	0.3	0.5
A3	1.1	0.2	0.3	0.0	0.0	0.0	0.0	0.1
A4	0.0	0.0	0.4	0.0	0.0	0.9	0.1	0.0
A5	0.4	0.2	0.0	0.0	0.0	0.4	0.0	0.3
A6	0.2	0.3	1.1	0.1	0.0	1.3	0.1	0.1

A7	0.1	0.2	0.4	0.0	0.0	0.0	0.0	0.0
A'8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A'9	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.1
A10	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.1
A11	0.4	0.5	0.6	0.0	0.2	0.5	0.3	0.3
A12	0.0	0.4	0.1	0.0	0.0	2.1	0.0	0.0
A14	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.1
A'16	0.0	1.3	1.4	0.1	0.2	0.2	0.0	0.4
A'17	0.5	0.8	1.1	1.9	0.4	0.4	1.2	1.0
A19	0.8	0.0	0.0	0.1	0.1	1.0	0.2	0.4
ND A'1	1.5	1.5	1.8	1.1	1.7	1.0	1.2	0.7
ND A'2	0.0	0.0	0.0	0.0	0.2	0.4	0.0	0.0
ND A'3	0.0	0.0	0.4	0.0	0.2	0.0	0.1	0.0
ND A'4	0.4	0.0	0.0	0.3	0.3	0.0	0.0	0.1
ND A'5	0.5	0.0	0.2	0.0	0.3	0.0	0.0	0.1
ND A'6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table D3. Population percentage of putative ecotypes (PEs) over a 5 day period after ~92% reduction in downwelling photon irradiance (MD). A and B indicate replicate samples where applicable.

PE	M D 0A	M D 0B	MD 1A	MD 1B	MD 2B	MD 2A	MD 3A	MD 3B	MD 4A	MD 5A	MD 5B
B′1	1.6	0.0	3.9	1.2	3.1	3.2	1.4	0.5	0.3	0.0	1.1
B′2	16.9	4.2	8.6	1.3	9.3	14.9	7.4	4.1	7.7	21.0	6.6
B′3	2.8	8.4	5.1	0.5	2.8	1.4	2.2	7.7	5.9	0.0	5.3
B′5	1.8	0.4	7.3	0.5	2.2	1.3	1.7	2.9	3.6	2.6	5.1
B′7	2.1	4.6	3.7	0.9	3.0	4.1	2.6	1.5	2.4	4.2	5.0
B′8	5.3	4.1	6.3	1.1	8.5	7.7	18.5	4.2	10.1	12.4	7.4
B′9	1.4	0.7	3.1	7.4	6.0	1.7	3.1	4.4	2.9	1.2	2.4
B′11	2.9	0.5	4.1	0.5	6.6	5.4	7.4	2.3	7.9	0.3	3.7
B′12	3.1	3.0	14.6	3.2	9.9	7.8	9.5	19.1	12.3	6.5	12.5
ND B′1	0.2	0.0	1.2	0.3	0.6	1.8	0.3	0.5	0.5	0.0	0.6
ND B′2	0.3	0.0	1.4	0.0	0.6	1.0	2.1	0.9	2.8	0.0	0.4
ND B′3	4.8	5.7	2.7	2.5	3.7	5.7	3.0	5.3	5.2	0.0	3.8
ND B′4	4.5	0.0	4.9	0.9	3.1	3.6	2.4	2.8	2.9	0.0	3.7
ND B′5	6.4	0.0	9.4	2.2	9.1	6.8	13.0	7.1	8.0	0.0	9.9
ND B′6	2.2	0.0	5.6	0.5	3.7	4.1	2.9	6.9	5.8	0.0	3.4
ND B′7	6.1	16.3	4.6	2.6	4.2	5.5	8.0	8.3	8.6	14.2	7.0
ND B′8	0.1	0.0	2.0	0.2	2.8	4.5	3.5	0.6	1.0	0.0	2.4
ND B′9	0.0	0.0	0.5	0.2	1.1	0.7	2.1	0.0	0.0	0.0	1.1
A1	1.0	3.1	0.6	2.7	0.6	0.3	0.2	0.0	0.5	0.7	1.4
A2	0.1	1.7	0.4	0.4	0.5	0.2	0.0	0.0	0.0	0.7	0.7
A3	0.5	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.1
A4	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.0
A5	0.4	0.0	0.3	0.7	0.3	0.0	0.0	0.0	0.0	0.0	0.0

A6	0.6	1.7	0.0	0.4	0.2	0.1	0.0	0.0	0.0	0.6	0.2
A7	0.3	0.0	0.2	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.7
A’8	0.0	0.0	0.0	0.7	0.0	0.1	0.0	0.2	0.0	0.0	0.0
A’9	6.7	9.5	0.0	0.3	0.3	1.0	0.1	3.0	0.1	5.2	0.3
A10	1.8	1.0	0.3	10.9	0.1	0.6	0.2	0.4	0.0	0.3	0.7
A11	0.1	0.0	0.0	0.9	0.4	0.0	0.0	0.0	0.0	0.0	0.3
A12	1.2	10.5	0.0	0.4	0.0	0.4	0.1	0.0	0.0	2.6	0.0
A14	0.4	0.2	0.3	0.0	0.3	0.2	0.0	0.0	0.0	0.0	0.5
A’16	5.7	6.0	0.1	6.9	0.8	1.0	0.0	3.8	0.4	0.5	0.9
A’17	5.2	7.7	0.4	1.2	0.1	1.3	0.2	1.5	0.2	15.6	0.3
A19	1.3	4.9	2.9	0.0	2.8	2.7	1.6	2.0	3.3	0.0	1.2
ND A’1	5.4	0.0	1.0	0.9	0.9	1.4	1.4	0.7	0.6	5.1	1.2
ND A’2	0.1	0.0	0.0	8.2	0.0	1.0	0.3	0.4	0.0	0.0	1.2
ND A’3	0.8	1.6	0.0	2.5	0.0	0.1	0.3	0.0	0.2	0.0	0.3
ND A’4	1.8	1.6	0.3	12.2	0.7	1.8	0.7	3.7	0.2	1.8	1.6
ND A’5	0.0	2.4	0.8	16.7	1.6	0.0	0.7	0.0	0.0	1.1	2.0
ND A’6	1.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table D4. Population percentages of putative ecotypes (PEs) over a 5 day period after exclusion of UV-light (MUV). A and B indicate replicate samples where applicable.

PE	M UV 0A	M UV 0B	MUV1A	MUV1B	MUV2A	MUV2B	MUV3A	MUV3B	MUV4A	MUV4B	MUV5A	MUV5B
B′1	3.1	1.4	3.6	3.5	3.7	4.6	3.3	1.7	2.6	1.2	4.1	1.1
B′2	6.7	7.1	7.5	2.5	5.2	4.2	7.2	5.0	8.7	7.8	7.7	5.0
B′3	4.7	2.6	1.7	1.4	2.2	2.9	3.9	0.5	3.9	1.6	4.0	7.0
B′5	1.6	2.5	2.5	2.6	1.6	2.2	3.1	4.7	3.1	2.2	3.5	2.8
B′7	4.0	6.2	4.8	3.5	4.9	5.1	5.8	4.2	5.2	5.7	3.3	1.7
B′8	5.3	10.0	6.5	11.0	9.4	5.7	15.7	15.9	11.8	6.0	6.8	8.2
B′9	1.4	4.8	12.3	14.3	17.1	15.5	2.1	2.1	1.3	3.6	2.0	0.5
B′11	2.2	5.5	5.5	5.2	3.9	7.0	9.3	7.6	5.4	8.3	3.6	3.7
B′12	11.0	11.3	8.1	9.9	12.2	9.4	12.2	9.5	18.7	15.9	13.0	20.7
ND B′1	0.6	0.9	5.4	2.2	1.9	2.9	1.2	0.1	1.3	1.2	0.8	1.0
ND B′2	0.0	0.7	5.6	2.8	3.1	4.6	0.8	0.2	0.9	0.5	0.2	0.7
ND B′3	2.5	7.3	7.3	4.6	4.6	7.0	4.3	3.0	3.4	4.7	4.5	5.1
ND B′4	3.0	3.4	13.9	12.4	8.7	8.8	3.7	2.6	2.8	2.5	5.1	2.3
ND B′5	11.9	10.2	4.5	4.9	6.1	4.9	10.0	18.4	8.6	8.3	7.5	9.1
ND B′6	4.4	7.2	2.6	3.8	5.9	5.6	5.1	8.1	6.9	9.6	6.4	7.7
ND B′7	11.4	6.8	2.6	3.1	3.0	2.7	3.2	9.2	3.8	7.6	8.8	10.9
ND B′8	0.1	0.4	2.0	1.8	1.7	2.6	2.4	1.6	2.8	1.4	0.2	0.7
ND B′9	0.0	0.0	0.3	0.0	0.5	0.3	0.6	0.0	0.3	0.0	0.0	0.0
A1	2.1	0.8	0.0	0.0	0.0	0.0	0.4	0.6	0.1	0.4	0.3	0.0
A2	0.9	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0
A3	0.2	0.4	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
A4	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
A5	1.1	0.2	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0
A6	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.1	0.0

A7	0.2	0.2	0.0	0.0	0.0	0.0	0.3	0.0	0.1	0.0	0.0	0.0
A’8	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A’9	0.7	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	1.1
A10	0.6	0.6	0.0	0.0	0.0	0.0	0.1	0.0	0.2	0.0	0.4	0.0
A11	0.1	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A12	1.4	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.1
A14	0.0	0.6	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.2	0.0	0.0
A’16	0.7	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.8	0.7
A’17	4.4	0.6	0.1	0.1	0.2	0.4	0.2	0.9	0.3	0.2	0.7	0.5
A19	5.7	0.6	0.0	2.3	0.8	1.6	0.5	0.0	2.0	1.6	5.0	1.4
ND A’1	7.1	1.7	0.2	0.5	0.4	0.5	0.8	1.9	0.9	1.3	1.9	2.7
ND A’2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.9	0.2	0.0
ND A’3	0.1	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0
ND A’4	0.7	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.5	0.8	1.5	2.1
ND A’5	0.0	0.0	0.0	0.2	0.2	0.0	0.2	0.0	0.3	0.0	0.0	0.4
ND A’6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table D5. Putative ecotype (PE) population percentages in replicate samples over a 5 day period (0-5) after removal of the top green layer (MS). A (left) and B (right) indicate replicate samples where applicable.

PE	MS0A	MS0B	MS 1A	MS1B	MS2A	MS3A	MS4A	MS4B	MS5A	MS5B
B′1	3.11	0.67	4.18	0.53	0.22	1.58	2.74	1.01	0.00	0.25
B′2	2.66	6.28	3.40	0.70	1.53	5.78	0.44	2.15	0.34	0.13
B′3	5.01	5.60	2.11	0.47	0.11	0.00	7.07	2.90	0.25	0.13
B′5	2.89	1.84	0.43	0.00	0.11	0.00	5.04	0.57	0.00	0.00
B′7	4.71	2.82	10.98	0.76	0.38	4.94	2.56	2.53	0.34	0.00
B′8	11.47	7.17	7.46	2.45	0.49	0.42	1.50	3.28	1.01	0.44
B′9	0.84	2.78	5.08	3.45	1.92	32.07	0.62	1.01	1.68	1.39
B′11	3.04	0.85	0.94	0.00	0.11	0.00	2.48	0.38	0.25	0.00
B′12	6.15	8.65	10.20	1.29	0.71	18.93	3.01	7.26	0.92	0.44
ND B′1	0.15	0.09	0.00	0.29	0.00	0.00	0.18	0.00	0.00	0.00
ND B′2	0.00	0.00	0.20	0.00	0.00	0.00	0.18	0.00	0.00	0.00
ND B′3	1.90	4.75	2.19	0.29	0.00	0.00	0.00	0.13	0.84	0.00
ND B′4	1.67	4.93	1.05	0.00	0.22	0.00	0.00	1.14	0.00	0.00
ND B′5	5.62	17.35	12.42	1.81	0.66	2.73	0.62	4.36	0.84	0.38
ND B′6	2.81	3.32	1.99	0.00	0.49	0.00	2.65	2.97	0.84	0.25
ND B′7	3.95	17.93	10.78	1.87	0.71	0.00	0.53	5.24	0.42	1.01
ND B′8	0.00	0.27	0.63	0.00	0.00	0.00	0.80	0.57	0.00	0.00
ND B′9	0.00	0.00	0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A1	4.78	2.11	3.13	1.23	1.70	0.00	0.18	1.70	2.52	2.59
A2	6.23	1.12	0.35	0.41	0.44	0.00	5.66	0.32	0.42	0.57
A3	0.91	0.58	3.13	1.34	1.86	0.00	0.00	1.07	2.43	2.40
A4	0.00	0.00	0.00	0.00	0.00	0.00	2.03	0.00	0.00	0.00
A6	0.76	0.00	0.16	0.23	0.66	0.00	13.00	0.44	0.92	0.32
A7	3.64	0.54	0.20	0.76	0.27	0.00	4.07	0.95	0.76	0.95

A14	0.00	0.18	0.12	0.18	0.00	0.00	2.03	0.00	0.34	0.06
ND A1	0.68	0.99	2.77	0.53	0.88	0.00	0.00	3.16	0.00	0.00
ND A2	1.29	0.63	1.17	3.68	4.11	0.00	2.03	4.42	8.64	7.00
ND A3	3.42	0.00	0.31	0.53	0.99	0.00	3.71	1.07	1.43	1.96
ND A4	0.68	0.63	0.82	0.12	0.60	0.00	1.95	0.82	2.27	2.59
ND A5	3.04	1.08	0.27	1.11	0.93	0.00	5.57	1.33	1.93	1.77
A’8	0.08	0.00	0.27	2.28	3.34	0.00	0.27	2.02	6.63	7.13
A’9	1.67	0.81	0.00	8.29	9.26	0.63	0.88	6.50	5.62	7.76
A’16	1.06	1.12	0.78	11.97	9.36	0.00	0.18	3.72	7.30	6.37
ND A’1	1.06	1.79	0.86	0.47	0.16	0.00	0.00	0.25	0.00	0.00
ND A’2	0.38	0.54	0.51	4.03	5.42	0.00	2.83	5.49	10.40	12.24
ND A’3	0.61	0.09	0.08	3.21	3.01	0.00	0.62	2.84	4.70	4.42
ND A’4	0.61	0.81	1.72	18.11	20.76	24.19	0.62	5.87	15.35	16.91
ND A’5	0.08	0.00	1.80	1.93	2.14	0.00	0.00	3.41	1.09	0.00
ND A’6	0.00	0.00	0.00	0.29	0.11	0.00	0.00	0.00	0.17	0.32
ND A’7	1.06	0.58	1.45	21.61	19.93	8.73	0.80	6.63	13.00	10.66

Table D6. Putative ecotype (PE) population percentages in replicate samples over a 5 day period (0-5) after removal of the top green layer and ~92% reduction in downwelling photon irradiance (MSD). A (left) and B (right) indicate replicate samples where applicable.

PE	MSD 0A	MSD 0B	MSSD A	MSSD B	MSD 1A	MSD 1B	MSD 2A	MSD 2B	MSD 3A	MS D 3B	MSD 4A	MSD 4B	MSD 5A
B′1	4.99	5.17	2.82	3.33	1.00	1.04	0.75	1.55	1.50	1.33	0.37	0.00	2.02
B′2	9.46	8.59	9.31	12.11	7.28	3.44	1.16	3.38	1.79	2.00	0.97	1.35	2.39
B′3	3.47	5.12	3.35	6.11	4.44	4.02	0.64	0.88	1.43	1.52	1.21	1.54	3.25
B′5	2.26	3.62	1.76	2.06	2.22	1.69	1.39	1.62	1.57	1.03	0.88	1.28	1.04
B′7	3.00	5.17	6.49	2.00	4.90	2.27	1.62	1.35	1.43	2.18	0.65	1.48	3.31
B′8	7.04	8.09	18.20	6.33	7.97	4.28	0.64	1.82	0.36	2.43	0.42	0.77	1.35
B′9	2.57	2.96	1.38	8.17	4.52	0.78	4.10	0.14	5.65	11.95	7.05	3.73	4.29
B′11	4.78	4.52	4.95	3.50	6.21	1.30	0.52	2.30	0.72	1.27	0.60	1.86	1.53
B′12	12.93	11.95	8.14	8.78	10.80	6.29	2.83	4.80	1.79	2.97	1.67	1.99	1.53
ND B′1	0.74	1.00	1.22	0.17	1.69	0.45	0.17	0.07	0.00	0.00	0.00	0.00	0.12
ND B′2	2.57	0.75	1.12	0.89	0.00	0.13	0.00	0.00	0.00	0.30	0.00	0.19	0.12
ND B′3	5.20	4.67	2.82	2.78	5.13	1.36	0.64	1.89	0.72	0.91	0.46	0.83	1.04
ND B′4	6.10	6.73	4.15	7.50	5.29	1.95	0.98	1.89	0.79	0.49	0.51	1.03	0.31
ND B′5	10.30	8.19	9.58	5.83	8.97	5.97	2.14	8.38	2.07	3.52	2.55	0.96	2.02
ND B′6	4.83	5.42	2.93	4.56	4.98	3.24	0.40	2.43	1.79	0.73	1.11	1.03	0.61
ND B′7	4.68	4.02	4.47	8.61	8.74	3.37	1.21	7.16	1.79	1.27	1.85	1.73	0.86
ND B′8	2.57	1.61	2.98	0.33	0.69	0.00	0.23	0.14	0.00	0.49	0.00	0.00	0.43
ND B′9	1.42	0.40	0.75	0.00	0.00	3.89	0.00	0.20	0.00	0.00	0.00	0.00	0.12
A1	0.63	0.25	0.43	0.39	0.23	0.91	1.33	1.22	0.93	0.42	1.07	0.51	2.45
A2	0.79	0.40	0.80	0.33	0.00	0.00	0.00	0.54	0.29	0.12	0.14	0.13	1.47
A3	0.16	0.00	0.32	0.28	0.38	0.00	0.00	0.34	0.00	0.00	0.00	0.00	0.00
A4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00

A6	0.26	0.00	0.00	0.00	0.00	0.00	0.12	0.20	0.21	0.00	0.19	0.00	0.37
A7	0.16	0.20	0.16	0.11	0.00	0.52	0.00	0.68	0.21	0.24	0.00	0.32	0.86
A14	0.32	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.13	0.00
A'8	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.74	1.93	0.18	2.09	1.54	0.67
A'9	0.00	0.00	0.00	0.28	0.92	1.10	2.20	7.43	12.02	1.39	14.84	8.03	1.65
A'16	0.26	0.95	0.69	0.22	0.92	1.23	0.87	6.69	9.16	3.15	8.25	6.74	2.27
ND A1	2.42	1.36	0.96	10.44	1.23	0.84	4.22	4.26	2.50	1.27	2.09	0.00	2.33
ND A2	0.16	0.10	0.27	0.39	0.23	2.33	2.43	1.15	1.50	1.03	1.39	0.39	2.21
ND A3	0.00	0.15	0.27	0.00	0.00	0.32	0.81	0.47	0.50	0.24	0.00	0.26	0.49
ND A4	0.26	0.15	0.37	0.17	0.00	0.00	0.00	0.20	0.00	0.36	0.09	0.13	0.12
ND A5	0.11	0.00	0.00	0.11	0.15	0.00	0.00	0.61	0.36	0.00	0.56	0.00	0.43
ND A'1	1.68	2.56	2.66	0.78	1.15	0.78	0.69	1.22	0.57	1.52	0.00	0.96	0.25
ND A'2	0.00	0.00	0.00	0.72	0.00	1.23	3.12	0.95	1.43	1.27	1.25	1.41	0.74
ND A'3	0.00	0.65	0.00	0.50	0.00	1.36	1.33	0.88	2.00	1.52	1.53	1.93	5.94
ND A'4	0.42	0.20	0.11	0.22	2.15	16.99	18.55	8.99	18.17	16.31	17.66	23.06	10.72
ND A'5	0.21	0.10	0.96	0.00	0.00	12.78	29.02	2.09	2.22	21.22	1.95	2.38	29.72
ND A'6	0.00	0.00	0.00	0.00	0.31	0.71	0.00	0.41	0.07	0.00	0.00	0.13	1.23
ND A'7	0.11	0.25	0.00	0.00	0.77	7.52	10.35	11.89	16.17	8.55	18.03	22.09	6.99

Table D7. Putative ecotype (PE) population percentages in replicate samples over a 5 day period (0-5) after removal of the top green layer and exclusion of UV-light (MSUV). A (left) and B (right) indicate replicate samples where applicable.

PE	MSUV 0A	MSUV 0B	MSUV 1A	MSUV 1B	MSUV 2A	MSUV 2B	MSUV 3A	MSUV 3B	MSUV 4A	MSUV 4B	MSUV 5A	MSUV 5B
B′1	3.41	3.34	3.43	1.47	2.70	2.95	3.68	5.20	4.33	2.98	5.81	4.62
B′2	16.37	16.43	14.17	3.38	8.49	12.63	7.60	0.68	6.16	6.03	4.32	1.59
B′3	2.59	1.85	3.63	1.36	3.77	5.63	6.53	3.15	5.02	2.98	6.38	4.36
B′5	1.68	2.17	2.86	1.64	3.21	2.88	7.42	4.58	4.94	3.80	6.38	3.95
B′7	4.99	6.20	5.71	1.74	6.66	2.29	3.56	5.95	3.19	4.24	7.09	7.44
B′8	14.31	11.76	5.92	1.91	6.54	3.73	7.48	2.53	6.92	1.04	3.61	8.73
B′9	2.50	2.09	0.99	6.22	1.32	4.84	1.19	4.72	0.23	0.37	0.78	7.91
B′11	6.29	5.60	4.62	2.56	5.09	5.10	8.07	7.73	4.33	4.54	9.99	5.65
B′12	8.16	11.07	7.11	4.85	7.61	7.40	12.05	3.35	8.75	7.30	9.14	7.55
ND B′1	3.12	1.25	1.97	0.44	1.45	1.05	0.83	0.41	0.46	0.89	0.64	1.08
ND B′2	1.73	1.73	1.19	0.44	0.38	0.33	1.60	0.75	0.00	0.00	1.28	1.18
ND B′3	5.33	5.92	2.13	0.65	4.84	2.36	4.09	2.46	2.59	2.61	1.84	3.95
ND B′4	5.04	5.27	6.44	1.91	3.27	3.53	3.09	4.58	1.29	2.31	1.63	3.85
ND B′5	8.83	9.30	5.66	3.11	6.29	4.91	5.88	3.15	4.79	4.39	2.91	3.85
ND B′6	3.02	2.94	3.53	2.24	2.20	3.66	4.27	2.39	3.80	2.90	3.47	2.57
ND B′7	2.78	4.55	3.48	3.05	3.39	4.25	2.55	0.75	4.11	2.83	1.56	2.21
ND B′8	3.36	3.22	1.77	0.00	1.38	0.65	2.61	0.96	0.38	0.00	0.92	1.39
ND B′9	0.62	0.64	0.42	0.00	0.44	0.20	0.24	0.14	0.76	0.00	1.13	0.21
A1	0.14	0.08	0.67	1.15	0.00	0.59	0.30	15.32	5.78	0.30	0.28	0.21
A2	0.00	0.40	0.21	0.60	0.00	0.13	0.00	2.53	3.73	0.00	0.14	0.00
A3	0.00	0.32	0.00	0.22	0.13	0.00	0.00	0.00	0.15	0.15	0.00	0.00
A4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A6	0.00	0.00	0.00	0.11	0.00	0.00	0.00	1.09	0.76	0.00	0.35	0.00

A7	0.00	0.00	0.21	0.60	0.00	0.00	0.53	1.37	1.52	0.00	0.50	0.00
A14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00	0.00	0.00
ND A1	0.82	0.64	2.80	1.74	3.65	2.95	3.20	1.71	1.37	3.80	4.96	3.29
ND A2	0.00	0.00	0.78	3.60	0.44	0.92	0.00	1.78	1.75	0.89	0.85	0.77
ND A3	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.21	0.23	0.00	0.14	0.00
ND A4	0.00	0.00	0.10	0.38	0.13	0.26	0.18	2.87	1.22	0.00	0.28	0.21
ND A5	0.00	0.00	0.00	0.38	0.00	0.00	0.00	1.57	2.66	0.15	0.00	0.00
A’8	0.00	0.00	0.00	0.82	0.00	0.26	0.00	0.00	0.23	0.15	0.00	0.00
A’9	0.00	0.00	0.31	8.67	0.88	3.01	0.00	0.00	2.43	8.19	0.50	0.98
A’16	0.29	0.20	0.57	5.62	0.63	1.44	0.47	0.21	0.84	4.02	0.57	0.77
ND A’1	1.15	0.72	0.47	0.87	1.89	1.31	2.08	1.92	0.76	0.30	0.85	0.82
ND A’2	0.00	0.00	0.73	1.25	0.63	0.79	0.18	0.00	0.00	1.49	0.14	0.31
ND A’3	0.00	0.12	0.36	0.98	1.07	0.33	0.36	1.09	0.38	0.15	0.43	0.15
ND A’4	0.10	0.00	3.74	12.92	5.03	4.52	1.78	0.21	4.18	10.20	6.87	5.08
ND A’5	0.00	0.16	6.96	2.13	7.35	0.00	1.60	0.82	0.23	0.82	5.10	5.49
ND A’6	0.00	0.00	0.10	0.22	0.00	0.00	0.00	0.34	0.00	0.15	0.00	0.10
ND A’7	0.00	0.00	2.02	12.54	2.33	5.24	0.36	0.27	2.43	8.64	1.63	2.36

Table D8. Population percentage of putative ecotypes (PEs) over a 5 day period (0-5) in control (unaltered mat) high temperature samples (MTH). A (left) and B (right) indicate replicate samples where applicable. Missing days indicate failed sequencing reactions.

PE	MHT0A	M HT 0B	MHT1A	MHT2A	MHT2B	MHT3A	MHT4A	MHT4B	MHT5A
B′2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND B′3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND B′4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND B′5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND B′6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
B′8	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
B′5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′3	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
ND B′11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND B′12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′9	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0
ND B′15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND B′16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B′11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND B′18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND A1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND A4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND A5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

A6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND A8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND A9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND A’1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND A’2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND A’3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ND A’4	67.8	4.8	4.0	4.5	4.0	22.2	5.4	3.7	10.5
ND A’5	0.0	1.5	1.9	2.0	1.4	36.4	1.9	2.6	59.9
ND A’6	14.1	27.5	11.1	8.1	16.7	22.3	8.9	16.2	9.1
ND A’7	1.9	10.1	11.8	21.8	11.3	7.2	14.7	25.1	5.0
A’8	2.2	4.7	6.2	3.8	12.1	0.8	6.7	9.1	0.1
A’9	10.0	41.2	48.7	49.4	48.2	6.3	46.9	34.2	9.1
A’16	3.4	6.4	11.8	5.9	3.4	3.1	12.3	3.3	4.9

Table D9. Population percentage of putative ecotypes (PEs) over a 15 day period after shifting samples from ~58°C to ~67°C (M→HT). A and B indicate replicate samples where applicable. Missing days indicate failed sequencing reactions.

PE	MC 0A	M-->HT 1A	M-->HT 1B	M-->HT 2A	M-->HT 2B	M-->HT 3A	M-->HT 3B	M-->HT 4A	M-->HT 4B	M-->HT 15
B′1	8.8	2.6	0.9	2.7	1.2	4.5	0.0	2.0	1.0	0.0
B′2	4.7	12.6	13.6	5.2	13.5	6.2	9.9	6.4	9.7	0.0
B′3	3.0	1.8	2.9	1.1	1.0	2.4	0.0	1.5	0.6	0.0
B′5	2.2	2.6	0.7	0.0	1.3	3.3	0.0	1.9	0.8	0.0
B′7	3.6	4.3	3.4	12.6	3.8	12.5	30.5	8.4	3.1	0.0
B′8	9.9	15.4	13.4	16.4	9.0	10.8	0.0	18.3	12.8	0.0
B′9	2.1	2.3	7.8	1.7	4.2	1.9	5.8	3.2	1.6	0.0
B′11	3.3	4.2	6.0	8.5	9.6	4.7	2.7	3.1	7.1	0.0
B′12	11.6	6.6	7.0	14.3	4.5	12.8	0.0	11.1	4.4	0.0
ND B′1	1.2	0.4	2.2	0.0	0.3	1.3	0.0	0.8	0.3	0.0
ND B′2	1.6	1.1	0.4	0.0	1.0	0.2	0.9	0.8	0.3	0.0
ND B′3	2.6	3.9	8.6	3.1	3.4	3.9	0.0	5.7	3.3	0.0
ND B′4	2.1	3.1	6.7	0.0	4.9	2.3	0.0	3.5	3.0	0.0
ND B′5	6.1	6.9	4.5	0.0	10.5	6.1	0.0	9.7	11.7	0.0
ND B′6	3.7	6.2	7.7	0.0	3.0	3.5	0.0	4.4	2.6	0.0
ND B′7	3.9	3.6	3.8	16.6	12.1	3.8	9.9	3.5	12.5	0.0
ND B′8	2.0	1.6	0.5	0.0	0.7	1.4	1.3	2.3	0.4	0.0
ND B′9	3.1	0.5	0.0	2.3	0.0	0.9	0.0	3.0	0.4	0.0
A1	1.1	2.3	0.6	0.3	0.7	1.8	0.0	1.5	2.7	0.0
A2	0.6	0.4	0.0	0.0	0.1	0.5	0.0	0.0	0.8	0.0
A3	1.1	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0
A4	0.0	0.0	0.5	0.8	0.3	0.0	0.0	0.2	2.4	0.0
A5	0.4	0.4	0.0	0.0	0.1	0.3	37.2	0.5	0.5	0.0

A6	0.2	0.1	0.0	1.2	0.0	0.8	0.0	0.8	1.4	0.0
A7	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0
A’8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
A’9	0.0	0.0	0.0	0.5	1.4	0.4	0.0	0.0	1.2	0.0
A10	0.3	0.2	0.2	1.9	0.3	0.4	0.0	0.4	0.1	0.0
A11	0.4	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
A12	0.0	0.1	0.3	1.4	0.3	0.3	0.0	0.5	1.8	0.0
A14	0.3	0.8	0.2	0.4	0.1	0.4	0.0	0.5	1.4	0.0
A’16	0.0	0.1	0.0	0.2	0.5	0.1	0.0	0.4	0.8	0.0
A’17	0.5	0.4	0.3	0.3	1.3	0.5	0.0	0.6	0.9	0.0
A19	0.8	1.2	0.7	0.0	3.2	1.4	0.0	0.9	1.0	43.9
ND A’1	1.5	0.8	0.7	0.0	1.4	0.8	0.0	0.9	1.5	15.5
ND A’2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.7
ND A’3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.8
ND A’4	0.4	0.3	0.4	0.5	0.3	0.2	0.0	0.5	1.3	0.0
ND A’5	0.5	0.5	0.0	0.0	0.3	1.0	0.0	0.1	0.0	2.2
ND A’6	0.0	0.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0	6.4

Table D10. Mean average, standard deviation (SD) and percent SD of predominant putative ecotypes (PEs; ≥5% in a sample) of replicate samples in (A) scrapped, (B) reduced light and scrapped and (C) UV blocked and scrapped mat samples collected over of 5 day (D) period.

A	Top green layer removed (% STD range of 33-94)											
PE	**D0**	**SD**	**%SD**	**D1**	**SD**	**%SD**	**D4**	**SD**	**%SD**	**D5**	**SD**	**%SD**
NDB'5	9.3	3.0	32.6	5.0	3.5	71.4	2.4	1.3	52.6	0.7	0.4	55.2
B'8	11.5	8.3	72.2	7.1	7.5	105	2.5	2.6	106	0.6	0.3	53.5
B'9	7.4	1.8	23.9	5.7	6.3	110	5.1	3.0	58.6	0.7	0.3	49.9
B'12	1.8	1.4	76.1	4.3	1.2	27.1	0.8	0.3	34.0	1.5	0.2	13.4
A6	0.4	0.5	141	0.2	0.1	28.1	6.7	8.9	132	0.6	0.4	69.4
NDA'4	0.7	0.1	19.9	9.9	11.6	117	3.2	3.7	114	16.1	1.1	6.8
NDA'7	0.8	0.3	41.3	11.5	14.3	124	3.7	4.1	111	11.8	1.7	14.0
A'9	1.2	0.6	49.3	4.1	5.9	141	3.7	4.0	108	6.7	1.5	22.6
A'16	1.1	0.0	3.7	6.4	7.9	124	2.0	2.5	129	6.8	0.7	9.6

B	~92% light reduction and top green layer removed (% STD range of 29-77)																	
PE	**D0**	**SD**	**%SD**	**D0S**	**SD**	**%SD**	**D1**	**SD**	**%SD**	**D2**	**SD**	**%SD**	**D3**	**SD**	**%SD**	**D4**	**SD**	**%SD**
B'8	7.6	0.7	9.8	12.3	8.4	68.4	6.1	2.6	42.6	1.2	0.8	68.3	1.4	1.5	105	0.6	0.2	42.1
NDB'5	9.2	1.5	16.2	7.7	2.6	34.4	7.5	2.1	28.4	5.3	4.4	83.9	2.8	1.0	36.5	1.8	1.1	63.9
B'12	12	0.7	5.5	8.5	0.4	5.3	8.5	3.2	37.3	3.8	1.4	36.4	2.4	0.8	35.2	1.8	0.2	12.4
B'9	2.8	0.3	9.9	4.8	4.8	100	2.6	2.6	99.9	2.1	2.8	132	8.8	4.5	50.6	5.4	2.3	43.6
NDA'4	0.3	0.2	50.0	0.2	0.1	49.8	9.6	10.5	110	13.8	6.8	49.1	17.2	1.3	7.6	20.4	3.8	18.7
NDA'5	0.2	0.1	50.0	0.5	0.7	141	6.4	9.0	141	15.6	19.0	122	11.7	13.4	115	2.2	0.3	14.0
NDA'7	0.2	0.1	58.0	0.0	0.0	0	4.1	4.8	115	11.1	1.1	9.8	12.4	5.4	43.6	20.1	2.9	14.3
A'9	0.0	0.0	0	0.1	0.2	141	1.0	0.1	12.8	4.8	3.7	76.9	6.7	7.5	112	11.4	4.8	42.1
A'16	0.6	0.5	80.4	0.5	0.3	72.7	1.1	0.2	20.5	3.8	4.1	109	6.2	4.2	69.0	0.6	1.1	14.2

C	UV excluded and top green layer removed (% STD range of 40-95)																	
PE	**D0**	**SD**	**%SD**	**D1**	**SD**	**%SD**	**D2**	**SD**	**%SD**	**D3**	**SD**	**%SD**	**D4**	**SD**	**%SD**	**D5**	**SD**	**%SD**
B'2	16.4	0.0	0.2	8.8	7.6	87.0	10.6	2.9	27.8	4.1	4.9	118	6.1	0.1	1.5	3.0	1.9	65.3
B'12	13.0	1.8	13.8	3.9	2.8	72.5	5.1	2.0	38.7	5.0	3.5	69.9	4.0	4.2	104	6.2	3.6	58.6
B'8	9.6	2.1	21.4	6.0	1.6	26.7	7.5	0.1	2.0	7.7	6.1	79.9	8.0	1.0	12.8	8.3	1.1	13.5
ND	0.0	0.1	141	8.3	6.5	78.0	4.8	0.4	7.6	1.0	1.1	112	7.2	4.3	59.2	6.0	1.3	21.2
A'9	0.0	0.0	0	4.5	5.9	132	1.9	1.5	77.4	0.0	0.0	0	5.3	4.1	76.6	0.7	0.3	46.1

Table D11. Mean average, standard deviation (SD) and percent SD of predominant putative ecotypes (PEs; ≥5% in a sample) of replicate samples in (A) control, (B) reduced light and (C) UV blocked mat samples collected over a 5 day (D) period.

A	Control (% STD range of 35-53)								
PE	**D1**	**SD**	**%SD**	**D3**	**SD**	**%SD**	**D4**	**SD**	**%SD**
B′2	5.2	3.6	68.5	8.4	2.1	24.6	6.0	2.5	41.0
B′7	4.5	2.7	59.4	3.2	1.5	48.5	6.8	4.4	63.9
B′8	15.8	8.4	53.0	11.8	4.7	39.3	15.8	1.6	10.4
NDB′5	12	4.4	36.5	10.9	0.1	0.7	11.8	2.0	16.8
B′12	10.8	0.4	3.5	13.7	0.9	6.4	9.5	2.3	24.6
B′15	8.3	5.7	68.6	8.6	3.4	39.8	8.1	3.1	38.8
A1	1.4	1.1	78.5	1.8	0.3	85.3	1.8	1.9	104

B	~92% light reduction (% STD range of 19-84)														
PE	**D0 mean**	**SD**	**%SD**	**D1 mean**	**SD**	**%SD**	**D2 mean**	**SD**	**%SD**	**D3 mean**	**SD**	**%SD**	**D5 mean**	**SD**	**%SD**
B′2	10.6	9.0	84.9	5.0	5.2	104	12.1	4.0	32.9	5.7	2.3	40.5	13.8	10.2	74.0
B′8	4.7	0.8	18.1	3.7	3.7	99.9	8.1	0.6	7.1	11.3	10.1	89.4	9.9	3.5	35.7
ND B′5	3.2	4.6	141	5.8	5.1	88.2	8.0	1.6	20.6	10.0	4.2	41.6	5.0	7.0	141
B′12	3.0	0.1	2.6	8.9	8.1	91.1	8.9	1.5	16.8	14.3	6.7	47.2	9.9	4.8	48.1
B′15	11.2	7.2	64.2	3.6	1.4	38.8	4.8	0.9	18.3	8.2	0.3	3.1	10.6	5.1	48.3

C	UV exclusion (% STD range of 14-30)														
PE	**D1**	**SD**	**%SD**	**D2**	**SD**	**%SD**	**D3**	**SD**	**%SD**	**D4**	**SD**	**%SD**	**D5**	**SD**	**%SD**
B′2	5.0	3.5	71.3	4.7	0.7	14.3	6.1	1.5	24.9	8.2	0.7	7.9	6.4	1.9	30.3
ND B′4	13.1	1.0	7.8	8.7	0.1	1.0	3.1	0.8	25.0	2.6	0.2	8.7	3.7	2.0	53.4
B′8	8.8	3.2	36.9	7.6	2.6	34.6	15.8	0.1	0.8	8.9	4.1	45.6	7.5	1.0	13.2
ND B′5	4.7	0.3	6.5	5.5	0.8	15.1	14.2	5.9	41.8	8.4	0.2	2.3	8.3	1.1	13.3
B′9	13.3	1.4	10.5	16.3	1.2	7.1	2.1	0.0	0.5	2.5	1.7	67.1	1.3	1.0	80.5

APPENDIX E

SUPPLEMENTARY INFORMATION FOR CHAPTER 7

Section I: Problems with Multiplex Ti454-barcoding and Sequencing

Multiplex-sequencing is the simultaneous sequencing of multiple loci using different primers in a single reaction. In an attempt to maximize the number of loci and habitats analyzed in multi-locus sequence analyses, 3 loci (*rbsK*, *aroA* and *pcrA*) were sequenced simultaneously on one Ti454 plate containing 3 different sequence-specific primers. Since each primer and target locus has different chemistries and energetics of primer binding, primer bias was probably the reason for the resulting limited sequence output. Failed sequencing reactions are indicated in Appendix Table E1 (grey shading). *pcrA* was not successfully sequenced in any sample, and *aroA* was sequenced sporadically across habitats. *rbsK* was successfully sequenced in most samples, and had the highest numbers of sequences across all samples.

Appendix Tables E1. Matrix of 83 samples from habitats sequenced at 3 loci used in MLSA studies and *chp* (which was sequenced separately). White boxes indicate successful sequencing reactions. Grey boxes indicate failed sequencing reactions, possibly due to primer bias during multiplex-sequencing. Numbers in parentheses indicate replicate samples. *chp* was only sequenced for samples collected along the effluent flow path at ~1˚C intervals (55 to 65˚C). Black boxes indicate that *chp* was not sequenced for that sample.

	55°C	56°C	57°C	58°C	59°C	60°C	61°C	62°C
rbsK								
aroA								
pcrA								
chp								
	60°C	60°C V2(2)	60°C V3(2)	60°C	60°C V5(2)	60°C	60°C	60°C
rbsK								
aroA								
pcrA								
chp								
	60°C	63°C V1	63°C V2	63°C V3	63°C V4	63°C V5	63°C V6	63°C V7
rbsK								
aroA								
pcrA								
chp								
	63°C	63°C V7(2)	63°C V8(2)	63°C	63°C	65°C V1	65°C V2	65°C V3
rbsK								
aroA								
pcrA								
chp								
	65°C V12	65°C V1	65°C V2	65°C V3	65°C V4	65°C V5	65°C V6	65°C V7
rbsK								
aroA								
pcrA								
chp								
	LL Day	LL Day	LL Day					
rbsK								
aroA								
pcrA								
chp								

[a] "V"; 80 µm vertical mat subsections. Number indicates the position of subsections from the surface (1 to n).
[b] "LL"; samples collected after 92% light reduction 2, 4 and 6 days after shading.

	63°C	64°C	65°C	60°C V[a] 1	60°C V2	60°C V3	60°C V4	60°C V5
rbsK								
aroA								
pcrA								
chp								
	60°C V9(2)	60°C V10(2)	60°C V1(3)	60°C V2(3)	60°C V3(3)	60°C V4(3)	60°C V5(3)	60°C V6(3)
rbsK								
aroA								
pcrA								
chp								
	63°C V8	63°C V9	63°C V10	63°C V1(2)	63°C V2(2)	63°C V3(2)	63°C V4(2)	63°C V5(2)
rbsK								
aroA								
pcrA								
chp								
	65°C V4	65°C V5	65°C V6	65°C V7	65°C V8	65°C V9	65°C V10	65°C V11
rbsK								
aroA								
pcrA								
chp								
	65°C V8	65°C V9	65°C V10	LL[b] Day 0	LL Day 2	LL 4	LL Day 6	LL Day 0(2)
rbsK								
aroA								
pcrA								
chp								
rbsK								
aroA								
pcrA								
chp								

[a] "V"; 80 µm vertical mat subsections. Number indicates the position of subsections from the surface (1 to n).

[b] "LL"; samples collected after 92% light reduction 2, 4 and 6 days after shading.

APPENDIX F

SUPPLEMENTARY INFORMATION FOR CHAPTER 8

Appendix Table F1. Putative ecotype (PE) percent population and within-ecotype percentage of all dominant variant (DV) sequences and all high-frequency sequences (HFSs) within predominant A-like and B′-like PEs in the American Northwest (also see Chapter 4, Figures 4.4 and 4.5, pp. 99-103). HFSs (in bold) increase in relative abundance from left to right, and total to 100%. The percentages of low frequency sequences (LFSs) associated with each DV and HFS are also reported (numbers to the right of the HFS not in bold). Spring abbreviations are JS, Jack Stream; Per, Perpetual; Lev, Levee; Boz, Bozeman; LaD, La Duke; NM, New Mound; BLV, Bath Lake Vista; WE, White Elephant Back; CW, Clearwater; OS, Octopus; TBV, Twin Butte Vista; MS, Mushroom; Man, Mantrap; and HP, Heart Pool.

		Percent population of each HFS variant in a PE (LFS variation as a percent of the total HFS + associated LFS variation)													
PE B'12	%PE	**DV**	LFS	**HFS1**	LFS	**HFS2**	LFS	**HFS3**	LFS	**HFS4**	LFS	**HFS5**	LFS	**HFS6**	LFS
MS	11.3	**55.8**	21.6	**13.3**	12.6	**13.3**	20.6	**8.8**	4.4	**8.8**	7				
OS	11.5	**25.2**	16.8	**23.5**	7.8	**15.7**	15.8	**9.6**	15.3	**8.7**	4.8	**8.7**	12.7	**8.7**	20.1
TBV	4.3	**30.2**	23	**23.3**	14	**23.3**	22	**23.3**	24.1						
PE Boz B'1	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**	
MS	4.8	**52.1**	19.1	**27.1**	26.1	**20.8**									
Boz	37.3	**95.7**	17	**4.3**	11										
PE WT B'1	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**	
Man	1.2	**100**	12.4												
PE B'9	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**	
MS	4.4	**40.9**	23.7	**34.1**	18.2	**25**	17.9								
OS	5.5	**63.6**	21	**18.2**	23	**18.2**	13.7								
TBV	9	**50**	11.7	**27.8**	8.6	**11.1**	9.2	**11.1**	16						
LaD	20.9	**100**	18.4												
PE CW B'1	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**	
CW	13.8	**100**	23.9												

PE WT B'2	%PE	DV		HFS1		HFS2		HFS3		HFS4		HFS5		HFS6	
Man	25.2	**72.2**	13.1	**13.9**	25.8	**13.9**	21.1								
HP	2.1	**100**	14.2												

PE B'11	%PE	DV		HFS1		HFS2		HFS3		HFS4		HFS5		HFS6	
MS	1.6	**100**	16.7												
OS	2	**50**	19.9	**50**	13.4										
BLV	1.1	**100**	23.2												
CW	6.9	**100**	28.9												

PE LaD B'1	%PE	DV		HFS1		HFS2		HFS3		HFS4		HFS5		HFS6	
LaD	10.3	**82.5**	38.2	**17.5**	20.7										

PE ND B'3	%PE	DV		HFS1		HFS2		HFS3		HFS4		HFS5		HFS6	
MS	4.1	**26.8**	31.3	**24.4**	3.7	**24.4**	12.1	**24.4**	3.8						
OS	9.9	**35.4**	13.4	**34.3**	9.4	**10.1**	8.2	**10.1**	14.9	**10.1**	26				
TBV	2	**50**	11.9	**50**	13.5										
HP	2.5	**60**	20.6	**40**	12.8										
Boz	1	**100**	31.1												

PE B'5/6	%PE	DV		HFS1		HFS2		HFS3		HFS4		HFS5		HFS6	
MS	4.9	**32.7**	6.8	**26.5**	17.4	**20.4**	2.8	**20.4**	28.5						
OS	5.9	**32.2**	10.4	**16.9**	9.6	**16.9**	16.5	**16.9**	22.1	**16.9**	7.5				
Man	26.8	**100**	4.8												
HP	71.9	**22.1**	10.8	**77.9**	28.5										

PE B'7	%PE	DV		HFS1		HFS2		HFS3		HFS4		HFS5		HFS6	
MS	17.6	**40.3**	21.1	**18.8**	16.1	**10.2**	5.7	**9.7**	10.5	**8.5**	19	**6.8**	3.5	**5.7**	7.2

OS	26.6	**50.8**	4.8	**14.3**	2.3	**11.3**	12.3	**9.4**	17.8	**6.8**	8.5	**3.8**	12.8	**3.8**	9.5
TBV	4	**25**	13.8	**25**	16	**25**	34.7	**25**	24.1						
HP	6.5	**84.6**	40.2	**15.4**	17.6										

PE Man B1	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**	
WE	1	**100**	34.8												
BLV	3.1	**67.7**	13.2	**32.3**	11.4										
NM	3.9	**100**	31.1												
CW	5.7	**100**	27.5												

PE Man B2	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**	
MS	3.2	**68.8**	16.7	**31.3**	35.1										
WE	4	**100**	19.9												
BLV	71.5	**100**	23.2												
NM	1	**100**	28.9												
LaD	2.2	**100**	11.5												

PE Man B3	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**	
WE	3.8	**73.7**	9.1	**26.3**	7										
BLV	11.7	**74.4**	9.5	**8.5**	16.9	**8.5**	4.5	**8.5**	20						
NM	13.5	**71.9**	33.2	**11.1**	13.5	**9.6**	17.1	**7.4**	6.3						

PE Mam B4	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**	
WE	1	**100**	25.1												
BLV	4	**75**	16.8	**25**	29.5										
NM	2.9	**100**	19.4												

PE Mam B5	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**	
WE	95.7	**93.6**	26.5	**1.9**	14.8	**1.4**	2.2	**1**	8.9	**1**	4.5	**1**	14.1		

BLV	10.8	**79.6**	25.7	**11.1**	31.4	**9.3**	5.1							
NM	79.8	**96**	30.1	**1.5**	10.5	**1.3**	11.2	**1.3**	6					

PE ND B'4	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**
MS	4.2	**76.2**	20.5	**23.8**	3.8									

PE B'2	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**
MS	3.7	**73**	11.5	**27**	23.7									

PE B'8	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**
TBV	14.6	**55.5**	35.9	**17.8**	8.3	**11.6**	11.5	**8.2**	13.1	**6.8**	28.5			

PE NDB′5	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**
CW	5.3	**100**	34.9											
LaD	2.1	**100**	27.3											

PE NDB'`1	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**
MS	6.8	**35.3**	17	**20.6**	31.3	**14.7**	12.1	**14.7**	13.4	**14.7**	23.5			
OS	15.4	**42.9**	13.6	**22.1**	13.4	**11**	9	**11**	25.1	**6.5**	28.5	**6.5**	8.8	
TBV	6.5	**33.8**	8.5	**20**	11.9	**15.4**	19.5	**15.4**	17.2	**15.4**	27.1			
HP	11.6	**81.9**	19.8	**9.5**	20.6	**8.6**	16.2							
CW	6.6	**39.4**	28.7	**39.4**	31.1	**21.2**	6.9							
LaD	12.1	**70.2**	30.6	**29.8**	9.1									
Boz	41.4	**84.8**	22.5	**11.6**	17.4	**3.6**	2.5							

PE Boz B'2	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**
OS	1.2	**100**	32.1											
HP	10.5	**90.5**	25.9	**9.5**	4.2									
CW	9.3	**100**	21.9											

LaD	23.5	**50.6**	29.1	**49.4**	13.3											
Boz	17.9	**91.6**	26.6	**8.4**	9.9											
PE CW B'2	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**		
OS	1.2	**100**	9.7													
CW	27.2	**52.6**	8.6	**34.2**	19.5	**7.7**	15.9	**5.5**	8							
PE B'1	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**		
MS	6.5	**80**	31.6	**20**	24.9											
PE B'4	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**		
MS	3	**100**	34.1													
Ore PE A1	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**		
Lev	18.7	**40.6**	5.7	**28.3**	25.1	**16**	7.2	**8.6**	13.3	**6.4**	12.4					
Per	17.8	**47.8**	7.7	**22.5**	16.8	**15.2**	13.6	**9**	7.1	**5.6**	15.5					
Jacks	10.2	**76.5**	18.5	**13.7**	19.4	**9.8**	2.8									
Ore PE A2	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**		
Per	52.9	**55.8**	14.4	**13.8**	20.9	**11.2**	12.5	**9.6**	8.4	**9.6**	14.1					
Ore PE A3	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**		
Lev	79.6	**48.7**	26.6	**11.7**	34.1	**10.3**	6.8	**9.7**	6.4	**9.2**	11.3	**5.9**	18.2	**2.9**	5.8	
Jacks	54.1	**72.3**	30.5	**20.3**	12.5	**3.3**	14.9	**2.2**	5.6	**1.8**	5.9					
Ore PE A4	%PE	**DV**		**HFS1**		**HFS2**		**HFS3**		**HFS4**		**HFS5**		**HFS6**		
Lev	18	**43.3**	27.5	**24.4**	16.2	**13.3**	17.3	**13.3**	9.1	**5.6**	15.9					
Per	2.2	**54.5**	36.2	**45.5**	9.7											
Jacks	20.7	**52.7**	12.7	**36.7**	8.6	**5.8**	12.4	**4.8**	8.5							

Appendix Tables F2. Putative ecotype (PE) percent population and within-ecotype percent sequence total of dominant variants (DVs) and all high-frequency sequences (HFSs), within predominant B′-like PEs in Mushroom Spring over a period of years from 1996 to 2009 (also see Chapter 5, Figure 5.5, pp. 140). HFSs (in bold) increase in relative abundance from left to right, and total to 100%. The percentages of low frequency sequences (LFSs) associated with each DV and HFS are also reported (numbers to the right of the HFS not in bold).

1996

Percent population of each HFS variant in a PE (LFS variation as a percent of the total HFS + associated LFS variation)

PE	%PE	**DV**	LFS	**HFS 1**	LFS	**HFS2**	LFS	**HFS3**	LFS	**HFS 4**	LFS	**HFS5**	LFS	**HFS6**	LFS
B′1	10	**39**	18.4	**32**	28.3	**29**	18.2								
B′2	2.7	**100**	22.2												
ND B′1	3.9	**64.1**	5.8	**38.5**	63.2										
ND B′2	3.2	**68.8**	10	**31.3**	75										
B′7	17	**41.8**	10.2	**13.5**	17	**10.6**	11.5	**10**	16.7	**8.8**	35	**7.1**	9.5	**16.9**	6.7
B′9	4.5	**40**	29.6	**35.6**	7.7	**24.4**	0								
ND B′3	1.9	**100**	0												
B′12	19.2	**32.8**	8.9	**17.2**	33.7	**12**	25	**7.8**	5	**7.3**	39.1	**6.8**	7.7	**17.5**	38.5
B′8															
ND B′4															
ND B′5	1.6	**100**	57.1												

2006

Percent population of each HFS variant in a PE (and associated LFS variation)

PE	%PE	**DV**	LFS	**HFS1**	LFS	**HFS2**	LFS	**HFS3**	LFS
B′1	1.6	**100**	24						
B′2	11.9	**56.3**	9.3	**16**	6.7	**13.4**	60	**11.8**	13

ND B′1									
ND B′2	1	**100**	31.3						
B′7	4.3	**100**	14.5						
B′9	10.2	**46.3**	2.2	**28.8**	3.1	**25**	8.3	**1**	18.8
ND B′3									
B′12	9.3	**59.1**	10.2	**17.2**	29.4	**11.8**	44.4	**11.8**	22.2
B′8	9.6	**54.2**	7.3	**45.8**	17.1				
ND B′4	1	**100**	7.7						
ND B′5	2.9	**100**	33.8						

2008

Percent population of each HFS variant in a PE (and associated LFS variation)

PE	%PE	**DV**	LFS	**HFS1**	LFS	**HFS2**	LFS	**HFS3**	LFS	**HFS4**	LFS
B′1	1.4	**100**	13								
B′2	5.5	**100**	6.7								
ND B′1	1.5	**100**	24								
ND B′2											
B′7	2.7	**100**	6.8								
B′9	20.6	**45.1**	32.8	**25.7**	2.7	**14.1**	9.1	**10.2**	25	**4.9**	20
ND B′3											
B′12	4.8	**70**	10.1	**30**	11.8						
B′8	14.3	**62.9**	3.4	**28**	6.2	**9.1**	18.2				
ND B′4	5.2	**51.9**	29.5	**25**	19	**23.1**	15.8				
ND B′5	1.1	**100**	66.7								

2009											
Percent population of each HFS variant in a PE (and associated LFS variation)											
PE	%PE	**DV**	LFS	**HFS1**	LFS	**HFS 2**	LFS	**HFS 3**	LFS	**HFS4**	LFS
B′1	1.4	**100**	12.4								
B′2	24.8	**74.6**	2	**18.5**	8.4	**6.9**	6.6				
ND B′1	6.2	**40.3**	8.9	**21**	8.3	**21**	79.4	**17.7**	23		
ND B′2	3.4	**61.8**	10.4								
B′7	1.3	**100**	8.4								
B′9	20.6	**45.1**	3.9	**25.7**	9.7	**14.1**	6	**10.2**	3.4	**4.9**	4.8
ND B′3											
B′12	8.8	**79.5**	4	**20.5**	6.9						
B′8	1.5	**100**	4.2								
ND B′4											
ND B′5											