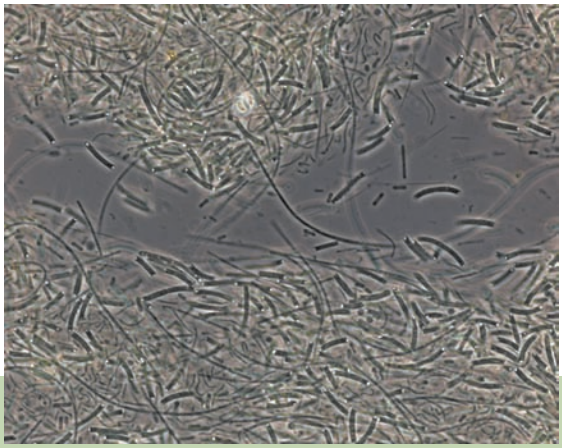
Microbial Diversity in Hot Spring Cyanobacterial Mats: Pattern and Prediction

David M. Ward[1*] | Frederick M. Cohan[2]

[1]Department of Land Resources and Environmental Sciences
Montana State University, Bozeman
[2]Department of Biology, Wesleyan University, Middletown, CT

Corresponding Author:
Land Resources and Environmental Sciences
334 Leon Johnson Hall
Montana State University
Bozeman, MT 59717-3120

Phone: 406.994.3401 Fax: 406.994.3933 E-mail: umbdw@montana.edu



ABSTRACT

Direct molecular analysis of the composition and structure of geothermal cyanobacterial mat communities has revealed diversity patterns suggesting that adaptive radiation and geographic isolation are the important drivers of cyanobacterial diversification. It is clear that 16S rRNA sequence variation cannot detect all ecological or geographic populations (ecotypes and geotypes, respectively), but it remains unclear what level of molecular resolution is required to do so. These empirical patterns are consistent with periodic selection theory for the evolution of prokaryotic populations, which predicts that adaptation and/or geographic isolation lead to terminal evolutionary clades with discrete ecological or geographic character. These clades, if long standing and not subject to geographic effects, may be detected as molecular sequence clusters to identify putative ecotypes within a community. An evolutionary simulation, called Star, is used to predict from the features of sequence data how to demarcate such sequence clusters. We are developing methods to determine whether the ecological diversity in mat communities is best predicted by periodic selection theory or by alternative theories based on rapid evolution of ephemeral ecotypes and/or the effects of geographic isolation.

Key Words

adaptive radiation
cyanobacteria
ecotype
geotype
internal transcribed spacer region
multi-locus sequence typing

1.0 INTRODUCTION

Since the mid-1980s, the Ward lab has been investigating the diversity of microorganisms inhabiting microbial mat communities found between ~50°C and 72°C (the upper temperature limit) in alkaline siliceous hot springs of the Lower Geyser Basin, Yellowstone National Park. Devoid of higher life forms, we consider these mats to be natural model communities in which to make fundamental insights into the composition, structure, and function of microbial communities. We began by studying the mat in Octopus Spring (**Figure 1A**), but due to a major hail damage event, switched our studies in the mid-1990s to Mushroom Spring, which we have found to be chemically (Ramsing and Ward, unpublished) and biologically (Ramsing et al. 2000; Ferris et al. 2003) similar. The mats are well laminated (**Figure 1B**), with an upper approximately 1-mm-thick green layer containing the main phototrophic microorganisms—unicellular cyanobacteria of the morphologically defined genus *Synechococcus,* and filamentous green nonsulfur-like bacteria (GNSLB), e.g., *Chloroflexus* and *Roseiflexus* (**Figure 1C**). Several features make these mats well suited for studies of microbial community ecology: (i) they are accessible; (ii) they have high biomass, thus enabling sophisticated molecular analysis; (iii) they are relatively stable systems; (iv) they exhibit well-defined environmental gradients (e.g., temperature, light) and are geographically separated; and (v) they have been studied for decades by Thomas D. Brock (Brock 1978), Richard Castenholz (Ward and Castenholz 2000), their students, and many other microbiologists, providing a base of information from which to launch detailed studies of microbial biodiversity and its significance. A final advantage appeared to be the simplicity of the mats as judged by traditional views of community composition, such as microscopy (**Figure 1C**) and cultivation (Ferris et al. 1996a). However, as the physicist Whitehead stated, "seek simplicity, but distrust it" (Begon et al. 1990). With the advent of molecular approaches for microbial biodiversity analysis, we were able to pass beyond the limitations of traditional methods. The purpose of this paper is to bring together a number of recently published key observations on molecular microbial diversity and its patterning relative to ecological and geographic gradients, and to evaluate the correspondence of these patterns to evolutionary theory, the focus of the Cohan lab (Cohan 2002d, 2004; Godreuil et al. 2005).

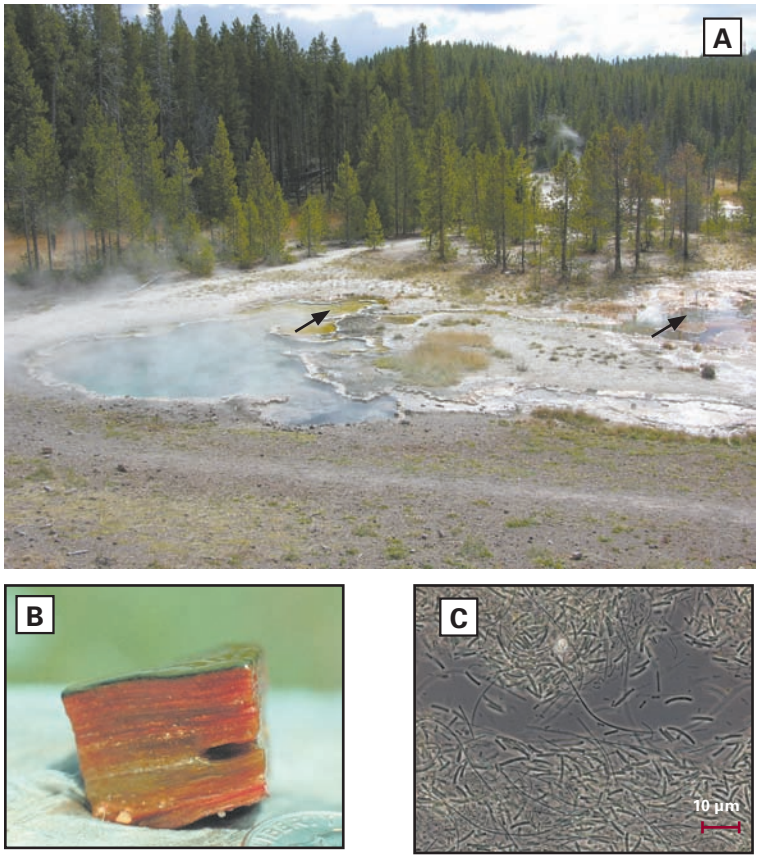


Figure 1. A typical Yellowstone alkaline siliceous hot spring microbial mat. **A.** Landscape view of Octopus Spring with arrows pointing to mat. **B.** Octopus Spring mat cross section at ~60°C. **C.** Photomicrograph of top 1-mm-thick green layer showing sausage-shaped *Synechococcus* and filamentous *Roseiflexus*. Bar indicates 10 µm.

2.0 METHODS

As the results presented herein have already been published, the reader is referred to the original papers for methodological details. We developed approaches to (i) clone the 16S rRNA genes directly from RNA and DNA extracted from mat communities (Ward et al. 1998), and (ii) study the distribution of the molecular variants we detected (Ferris et al 1996b). Distribution analysis was based on a gel separation method called denaturing gradient gel electrophoresis (DGGE), which separates PCR-amplified 16S rRNA gene segments on the basis of their melting characteristics, in turn a function of their sequence differences. Later we found it necessary to develop methods for studying the internal transcribed spacer (ITS) region of DNA separating 16S and 23S rRNA genes (Papke et al. 2003). Evolutionary simulation methods are described below (see 3.2 and 3.3).

3.0 RESULTS AND DISCUSSION

3.1 Molecular Variation and its Patterning

3.1.1 16S rRNA-based diversity. **Figure 2** illustrates the diversity of 16S rRNA sequences we have observed for cyanobacteria and GNSLB. Thick lines highlight unique cyanobacterial genotypes (sequence variants) found within the mat (**Figure 2A**), and boxes around clone designations highlight unique GNSLB genotypes (**Figure 2B**) against a background of representative diversity within each kingdom-level lineage. The total length of horizontal lines separating any two sequences in either tree is a quantitative measure of the genetic difference between these sequences. Clearly, more diversity was detected than meets the eye in a microscopic view (11 and 27 genotypes within single cyanobacterial and GNSLB morphotypes, respectively). The most readily detected genotypes (A/B lineage cyanobacteria and C lineage GNSLB) are distant relatives of readily cultivated cyanobacteria, *Synechococcus lividus* C1, and GNSLB, *Chloroflexus aurantiacus* (**Figure 2**). In fact, the genetic differences between cultivated and *in situ* cyanobacteria and GNSLB are on the order of the differences between ferns and flowering plants as measured by rRNA sequences! The most readily detected genotypes of cyanobacteria and GNSLB occur as clusters of closely

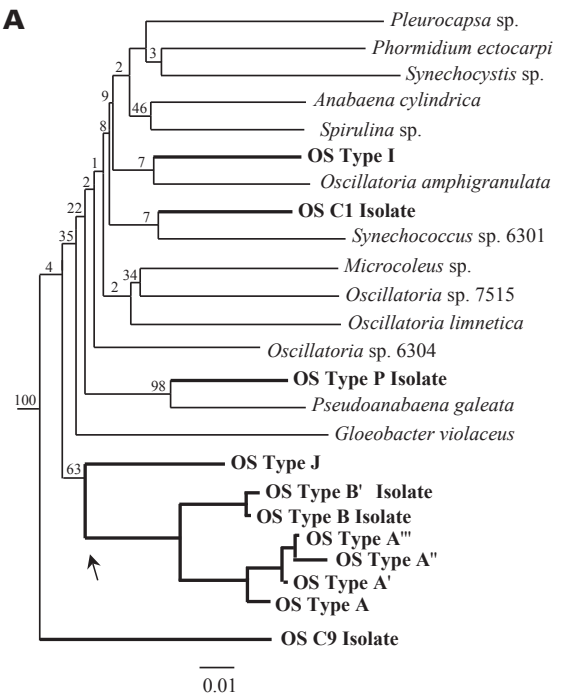


Figure 2A. Evolutionary history of 16S rRNA-based diversity of cyanobacteria (thick lines) detected in Octopus Spring against a background representative diversity within each kingdom-level lineage. Cyanobacterial tree, Ward et al. 1998.

related sequence types (e.g, A, A', A", B, and B' in the case of cyanobacteria). The C-like GNSLB sequences have been shown to be related to a filamentous bacteriochlorophyll *a*-rich organism, *Roseiflexus castenholzii*, recently cultivated from Japanese (Hanada et al. 2002) and Yellowstone (van der Meer et al., unpublished) hot springs.

While it is quite clear that the most readily detected genotypes are distantly related to known microorganisms (e.g., A/B *Synechococcus* vs. *S. lividus*, and type-C vs. *C. aurantiacus)*, the meaning of closely related sequence variants is much less clear. These might, for instance, simply represent sequence differences among operons (i.e., >1 variant per organism) or among ecologically interchangeable organisms. Clearly, there is a need to understand how molecular variation correlates with organismal variation. A first clue comes with the results

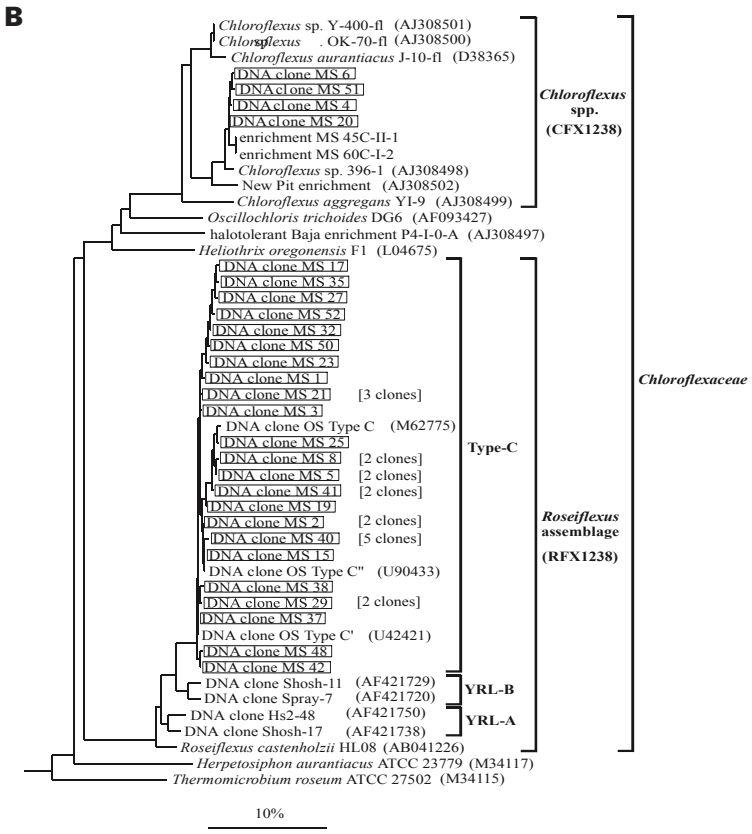


Figure 2B. Evolutionary history of 16S rRNA-based diversity of green nonsulfur-like bacteria, or GNSLB (boxed entries), detected in Mushroom Spring mats against a background representative diversity within each kingdom-level lineage. GNSLB tree, Nübel et al. 2002.

of distribution analysis. As shown in **Figure 3A (next page)**, closely related A/B-like cyanobacterial and C-like *Roseiflexus* genotypes are discretely distributed along a transect that runs parallel to flow. Since spring water cools as it flows away from the source, an obvious interpretation of the distribution results is that these closely related genotypes represent distinct populations with adaptations to different temperatures. A diversity of temperature-adapted populations had been previously suggested based on studies of *in situ* photosynthesis (Brock and Brock 1968; Bauld and Brock 1973) and pure-cultured *Synechococcus* strains (Peary and Castenholz 1964; Miller and Castenholz 2000), and we have recently verified this for Octopus Spring mat *Synechococcus* isolates of genotypes B, B', and A (Allewalt et al. submitted).

Similarly, as shown in **Figure 3B**, different A/B genotypes are found at different depths within the 1-mm-thick photic zone corresponding to differently pigmented *Synechococcus* populations that occur under varying light conditions (suggesting light-adapted strains). For instance, the pigment-rich *Synechococcus* population 400-700 μm below the mat surface, corresponding with the subsurface distribution of genotype A, receives only ~5% of the light available at the mat surface (Ferris et al. 2003). Of course temperature and/or light adaptations are not the only possible adaptations, as other parameters change with flow (e.g., nutrient type and amount; Ward and Castenholz 2000) and depth (e.g., oxygen, carbon dioxide, pH, hydrogen sulfide; Revsbech and Ward 1984). The implication of these distribution results is that the diversity of closely related A/B-like cyanobacterial and C-like *Roseiflexus* genotypes is the result of adaptive evolutionary radiation. That is, like Darwin's Galapagos finches or Hawaiian silversword plants (Ward et al. 2002), discrete ecologically specialized populations (ecotypes) are the result of the action of natural selection on genetic variation within populations. Another important point to emerge from distribution studies is that the ecological distinction of nearly identical variants (e.g., B and B' *Synechococcus* genotypes) calls into question whether the 16S rRNA locus varies sufficiently to detect all ecologically specialized populations (see 3.1.2).

We were also able to demonstrate that geographic isolation—the other major environmental determinant of divergence in plants and animals (Rosenzweig 1995)—affects divergence of hot spring cyanobacteria as well (Papke et al. 2003; **Figure 3C**). Genetically distinct cyanobacterial populations were found in different North American, Japanese, and New Zealand hot springs. For

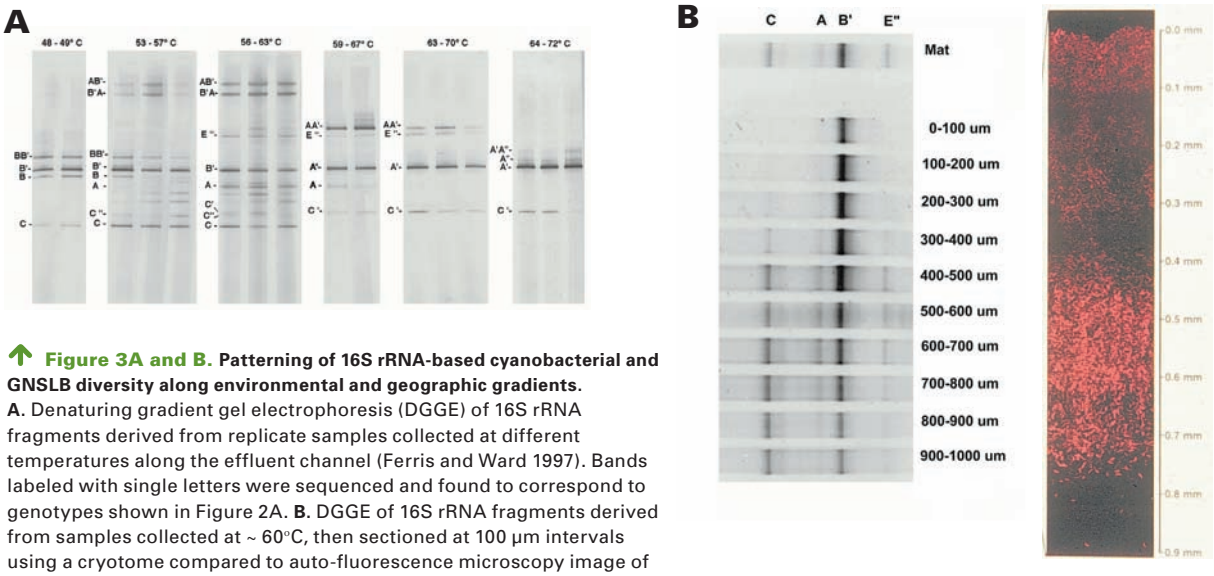


Figure 3A and B. Patterning of 16S rRNA-based cyanobacterial and GNSLB diversity along environmental and geographic gradients. **A.** Denaturing gradient gel electrophoresis (DGGE) of 16S rRNA fragments derived from replicate samples collected at different temperatures along the effluent channel (Ferris and Ward 1997). Bands labeled with single letters were sequenced and found to correspond to genotypes shown in Figure 2A. **B.** DGGE of 16S rRNA fragments derived from samples collected at ~ 60°C, then sectioned at 100 µm intervals using a cryotome compared to auto-fluorescence microscopy image of a vertical thin section of the same sample (Ramsing et al. 2000).

instance the A/B *Synechococcus* lineage was only found in North America; other cyanobacterial lineages were more widespread but variants from different countries formed distinct genetic clades. This biogeographical patterning could not be associated with environmental differences among the springs studied, suggesting that, like different flightless birds on different continents, physical isolation may act upon variation in microbial populations leading to the evolution of distinct geographic populations known as geotypes (Papke and Ward 2004). Very closely related genotypes appeared to exhibit unique geographic distributions (e.g., different sequences found in Oregon and the Greater Yellowstone Ecosystem; **Figure 3C**). This small degree of difference raises the question of whether the 16S rRNA locus exhibits enough sequence variation to detect all geotypes. Although it can be argued that hot springs are unusual among bacterial communities in their degree of geographic isolation (Fenchel 2003), the point has clearly been made that physical isolation as well as adaptation can affect the evolution of prokaryote diversity.

3.1.2 ITS-based diversity. The limitation of 16S rRNA for detecting cyanobacterial ecotypes and geotypes was investigated by studying a segment of the genome containing both an informative region of the 16S rRNA gene (i.e., that used to define 16S rRNA genotype) and the ITS region separating it from the 23S rRNA gene. The ITS region evolves more rapidly than the 16S rRNA gene, as it is under less evolutionary constraint. Thus, it was possible to explore genetic variation within defined 16S rRNA genotypes at higher resolution using the ITS locus. Ferris et al. (2003) reported that distinctly pigmented *Synechococcus* populations occurring at different depths in the 68°C Mushroom Spring mat have identical 16S rRNA sequences but ITS sequences that belong to distinct phylogenetic clades (**Figure 4A, page 192**). Interestingly, distinctly pigmented populations occurring at different depths in the 65°C mat showed no difference in 16S rRNA or ITS sequence (M.J. Ferris, unpublished). These either represent a single ecotype acclimated to different microenvironmental conditions, or more than one ecotype so closely related that yet higher resolution genetic markers are needed to discern them. Papke et al. (2003) reported that populations with identical 16S rRNA sequences but with distinct ITS sequences could be correlated with differences in geographic distribution within Yellowstone (**Figure 4B**) and Japan. Others have observed cases where populations with identical 16S rRNA sequences can only

be distinguished with higher resolution genetic markers (e.g., ITS, Rocap et al. 2002; protein-encoding genes, Palys et al. 1997, 2000). Such results not only confirmed our suspicion that the 16S rRNA locus is too conserved to detect all ecotypes and geotypes, but also called into question the ability of ITS and other genetic loci to detect these populations.

3.1.3 The importance of detecting ecotypes of a community. Clearly, it is important to detect each ecotype in microbial community ecology, as these are the populations that will occupy specific niches and respond uniquely in the changing environmental landscape. They can even be considered as species using an ecological species concept (Ward 1998). As Mayr (1982) said of species, they represent "the basic unit of ecology…no ecosystem can be fully understood until it has been dissected into its component species and until the mutual interactions of these species are understood." Hence, it is extremely important to understand the relationship between molecular variation and ecotype diversity. High-resolution methods for linking molecular diversity to ecotype diversity are needed.

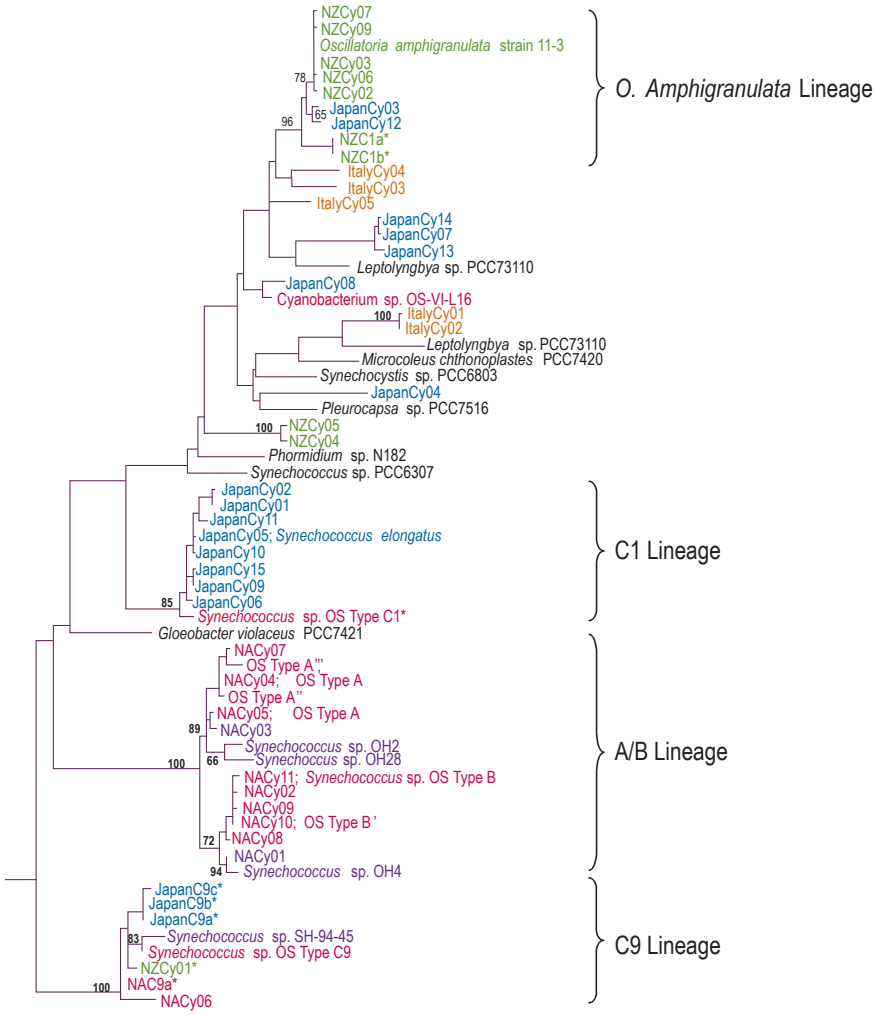


Figure 3C. Phylogenetic tree showing cyanobacterial 16S rRNA sequence variants detected in North America, red (except purple for Oregon); Japan, blue; New Zealand, green; and Italy, gold (Papke et al. 2003).

3.2 Molecular and Organismal Evolutionary Theory

Our aim is to characterize the full ecological and phylogenetic diversity of the *Synechococcus* populations inhabiting Yellowstone hot spring microbial mats. The extreme character of this habitat will no doubt make a full accounting of its biodiversity much simpler than in more mesic habitats. Nevertheless, a complete characterization of any community's diversity is only possible if organisms fall into discrete clusters of ecologically interchangeable organisms. Even in a simple community, a study of biodiversity would be complicated beyond feasibility if

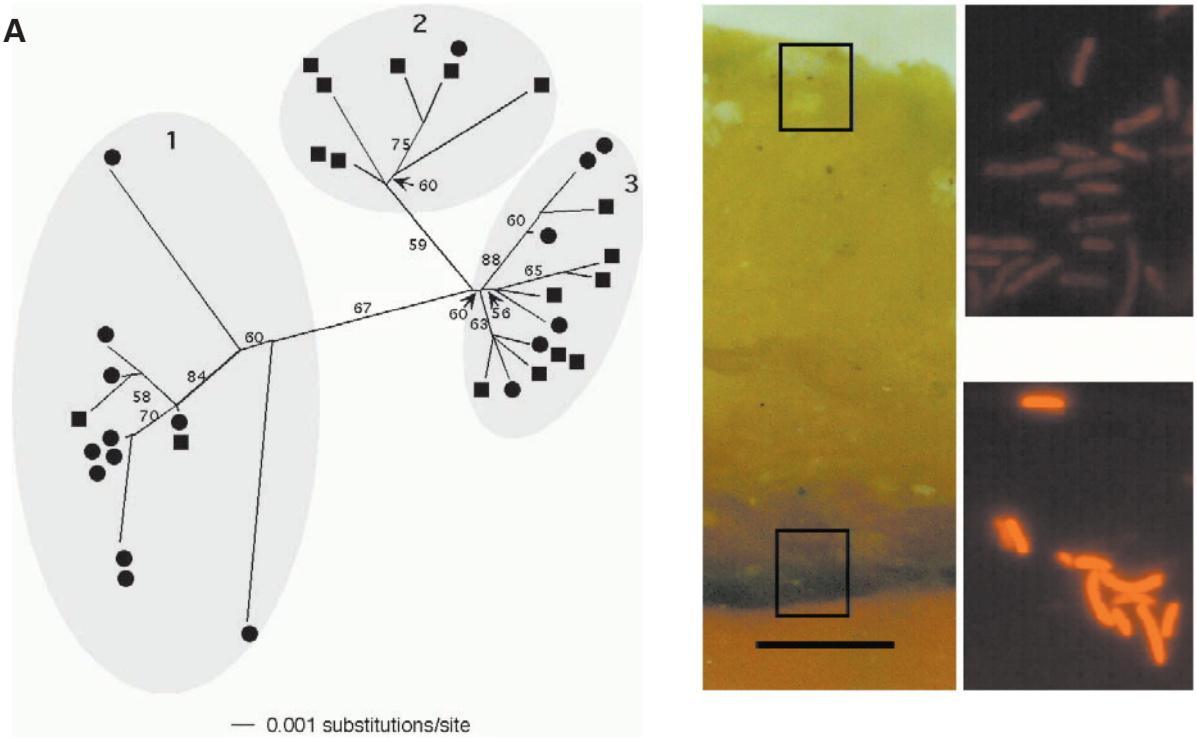


Figure 4. Hot spring cyanobacterial internal transcribed spacer (ITS) region-based diversity and its patterning.
A. Auto-fluorescence microscopy of the upper ~1 mm of a 68° Mushroom Spring mat sample showing differently pigmented *Synechococcus* populations at different depths and distinct ITS-defined phylogenetic clades showing genetic distinction of genotypes from the upper (circles) or lower (squares) populations (Ferris et al. 2003). **B.** ITS-based phylogeny of samples collected in different locations within Yellowstone National Park. Note that genotypes form clades that roughly correspond to northerly or southerly sampling sites (Papke et al. 2003).

nearly every individual organism in the community were ecologically unique. Fortunately, organisms from all walks of life (including bacteria, fungi, plants, and animals), and from every known community, seem to fall into discrete clusters of ecologically similar organisms (Claridge et al. 1997). We aim to investigate whether the organisms of Yellowstone hot spring microbial mats fall into a small number of ecologically distinct clusters. More profoundly, we will test whether these clusters assume the ecological and evolutionary dynamics of biological species, as understood by all modern species concepts (de Queiroz 1998).

3.2.1 Organisms inevitably fall into clusters. We first offer a theoretical rationale for why the bacteria in any community are expected to fall into discrete, ecologically distinct clusters. In the case of heterotrophs, individual organisms fall into clusters simply because their resources are packaged in discrete groups (Turelli et al. 2001). For example, one bacterial population may infect one host species while a closely related population infects another host species (as is the case for 11 sequence clusters within *Borrelia burgdorferi* sensu lato; Kurtenbach et al. 2002), or one population may infect one tissue of a particular host while another population infects another tissue of the same host, as is the case for skin-tropic and throat-tropic sequence clusters within *Streptococcus pyogenes* (McGregor et al. 2004). Similarly, many closely related populations of free-living heterotrophic bacteria appear to be ecologically distinct in the organic molecules they consume (Feldgarden et al. 2003). In all these cases, the resources are discrete and so are the bacteria consuming them.

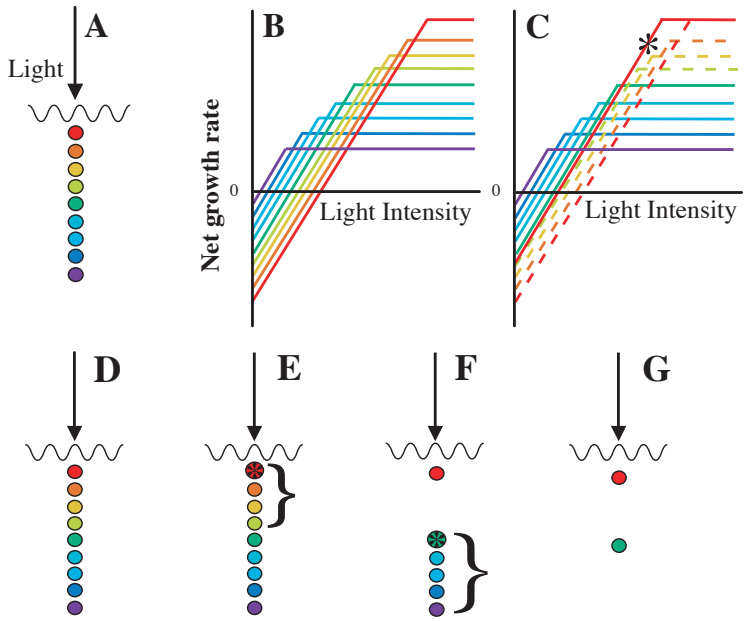


Figure 5. A model of ecological diversity in an aqueous habitat with a continuous gradient in light as a limiting resource. In this case, the limiting resource is a particular wavelength of light. **A.** The optimal depth and light level for each genotype (genotypes marked by colors). **B.** The growth rate of each genotype as a function of light intensity. Each genotype has the same linear response to increasing light, until it reaches the genotype's own saturation point. Note that each genotype is superior in growth rate over a narrow range of light levels. **C.** The bold red curve represents an adaptive mutant of the red genotype, which confers a broader range of optimal light intensity for growth. Note that now several of the genotypes adapted to slightly lower light levels no longer have a niche where they can successfully compete. **D-G.** An adaptive mutation (denoted by an asterisk) in the red genotype extinguishes three neighboring genotypes, and an adaptive mutation in the blue-green genotype extinguishes four neighboring genotypes. A discontinuous set of ecologically distinct populations results from inevitable improvements in efficiency.

Here we make the case that bacteria will fall into ecologically distinct clusters even when resources are not packaged in discrete units. We will illustrate our argument with a community of photosynthetic autotrophs in a hypothetical aquatic environment, in which we postulate that one particular wavelength of light is the only limiting resource. The light resource diminishes continuously in intensity, with the bright light at the surface steadily attenuated through shading and light scattering by bacteria. One might imagine that this continuous gradient in resources would yield a bacterial community with absolutely no clustering. For example, in **Figure 5A** there is a different organism optimally adapted to each level of light over the whole water column. Note that

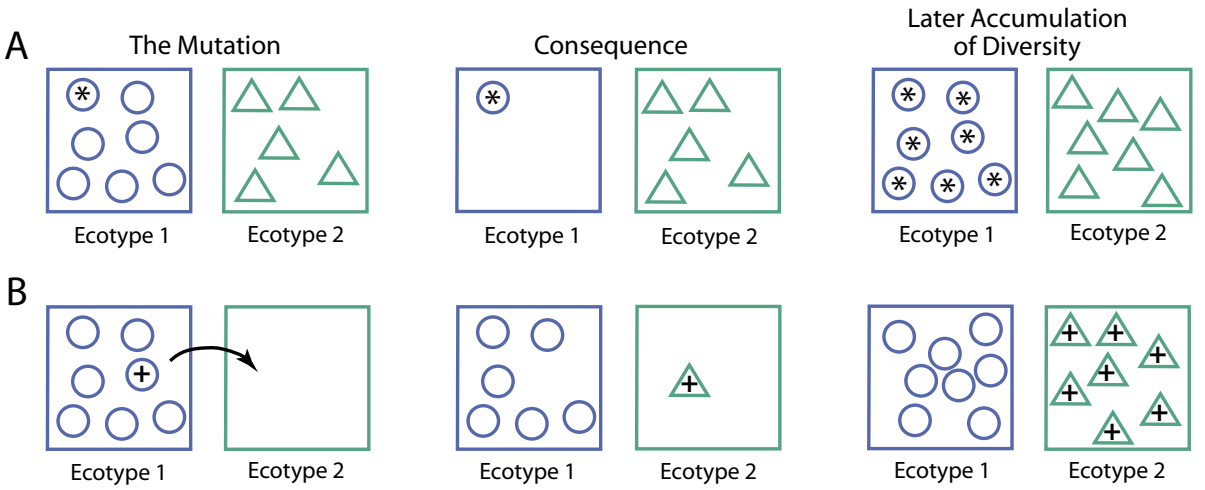


Figure 6. The effects of adaptive mutations on diversity within and between ecotypes. A. The effect of a periodic selection event. Here a mutant (or recombinant) with improved ability to compete for the resources of Ecotype 1, indicated by an asterisk, is able to extinguish the diversity within the same ecotype. The diversity within Ecotype 2 is not affected by periodic selection occurring within Ecotype 1. After the periodic selection, diversity once again accumulates within Ecotype 1. **B.** The effect of a niche-invasion mutation. Here a mutant, indicated by a plus sign, obtains the ability to utilize a new set of resources and thereby founds an ecotype (Cohan 2004a). Ecotype 2 begins as a clone, with no diversity, but eventually accumulates genetic diversity by mutation and recombination.

each organism can out-compete (i.e., grow faster than) all other organisms within a very narrow range of light intensity (**Figure 5B**). Barring evolution, and assuming a stable environment, infinitesimally different organisms could coexist indefinitely.

Nevertheless, we can demonstrate in a thought experiment that this lack of clustering is extremely unstable. Suppose that an organism adapted to grow faster in very intense light (e.g., the red organism in **Figure 5C**) should mutate, or acquire a gene by horizontal transfer, and thereby grow faster at lower light intensities as well (by becoming more efficient). Depending on the magnitude of the advantage of the adaptive mutation, the new mutant would out-compete a number of strains optimally adapted to lower light intensities at greater depths (**Figure 5C**). Subsequent mutations in other strains would also extinguish a range of ecological diversity (**Figure 5D-G**). What remains is a community with a small number of ecologically distinct populations (**Figure 5G**). Thus, even in an environment that is absolutely continuous in the resources available, it is inevitable that the organisms utilizing these resources will evolve into a set of discrete clusters representing ecologically distinct populations. We are currently developing a quantitative model demonstrating the relationship between the magnitude of adaptive mutations and the number of ecologically distinct populations that can coexist.

3.2.2 The diversity-purging effect of periodic selection. Next consider how we can use universally applicable molecular tools to discover the ecologically distinct populations of hot spring cyanobacterial mats and elsewhere. We begin by considering the fate of an adaptive mutation (e.g., conferring a higher growth rate) in one ecological population (**Figure 6A**). The mutant cell is expected to outgrow all other members of its population (consequence 1 in **Figure 6A**). Because of the rarity of recombination in bacteria (Cohan 2002b, 2002c; Feil et al. 1999, 2000, 2003; Maynard Smith et al. 1993), the nearly clonal descendants of the adaptive mutant replace the diversity within the population. Variants within an ecotype are ecologically interchangeable, and until an adaptive mutation favors one over the rest, any variant has an equivalent potential to carry the lineage forward

at the next periodic selection event. Thus, periodic selection favoring the adaptive mutant causes a purging of diversity within the population, at all loci within the genome (Atwood et al. 1951; Cohan 1994b, 2005; Koch 1974; Levin 1981). Within an ecologically homogeneous population, diversity is thus ephemeral, awaiting its demise with the next periodic selection event (Cohan 2002a).

How does divergence among bacteria become permanent? A mutation or a horizontal transfer event can allow a bacterium to invade a new ecological niche, such that it can consume a new set of resources or live in different environmental conditions such as different temperature or light conditions, the presence of an antibiotic, etc. (Cohan 1994a, 2001, 2005; **Figure 6B**). This niche-invading cell and its descendants will now form a new population, and their ecological distinctness will allow them to escape extinction caused by adaptive mutations of the former population. Note in **Figure 6A** that an adaptive mutant from one population purges the diversity among that population's membership, but diversity within the other population is untouched (Cohan 1994a, 2005). Populations may also diverge due to physical isolation (e.g., geographic isolation) because a periodic selection event in the portion of the population in one location does not affect the diversity in the portion of the population found in some distant place. We will consider geographical isolation further below.

To pursue further the evolutionary implications of ecological divergence, we must use a more precise definition of ecotype that includes the effects of periodic selection. Here an ecotype is defined as the set of organisms utilizing one ecological niche such that an adaptive mutant out-competes to extinction members of its own ecotype (and thereby purges genetic diversity within the ecotype, genome-wide); however, an adaptive mutant cannot out-compete members of other ecotypes owing to their ecological distinctness (Cohan 1994a; **Figure 6A**). Thus, the diversity within an ecotype is recurrently purged at all loci by periodic selection, but the divergence between ecotypes is not affected.

Ecotypes, as defined thus, have been shown to have all the quintessential qualities of species (Cohan 2001), as understood by evolutionary systematists (de Queiroz 1998). Despite some differences in emphasis among modern species concepts, species are nearly universally understood to be cohesive groups (de Queiroz 1998; Meglitsch 1954; Templeton 1989). That is, diversity within a species is constrained by some force of evolution. In the case of the highly sexual animals and plants, divergence within species is limited by the ability to exchange genes (Futuyma 1987; Mayr 1963). In the case of rarely sexual organisms, such as bacteria, genetic exchange between populations is never sufficient to prevent adaptive divergence (Cohan 1994a), so genetic exchange is not a cohesive force. Nevertheless, the very rarity of genetic exchange actually amplifies the cohesive power of natural selection (Cohan 1994a, 2001, 2002a; Templeton 1989). Under rare genetic exchange, the natural selection favoring an adaptive mutation purges the diversity not just at the locus under selection but at all loci.

In all modern species concepts, species also hold the property of being irreversibly separate (de Queiroz 1998; Simpson 1961; Wiley 1978). That is, while the organisms within a species are constrained from diverging by a force of cohesion (be it genetic exchange or periodic selection), organisms from different species are not so constrained. Periodic selection no longer puts a cap on the divergence between ecotypes, since the purging of diversity by periodic selection is confined (by definition) to a single ecotype. The ecological distinctness of ecotypes protects each ecotype from being extinguished by events that purge diversity within another ecotype.

Finally, all modern concepts of species hold that species should be ecologically distinct. This requirement comes from our understanding that no two species can coexist for long (whether sexual or asexual) unless they utilize different sets of resources, or utilize them best under different environmental conditions (Van Valen 1976). Bacterial ecotypes clearly fit this requirement of species-hood.

3.3 Sequence-based Discovery of Bacterial Ecotypes

3.3.1 Predictions of the Stable Ecotype Model. We consider first a Stable Ecotype Model, in which ecotypes are longstanding in their ecological coexistence and there

is no geographical isolation among the populations of an ecotype. If this model is correct, we should be able to discover all the ecotypes within a community through universal, sequence-based approaches. This point is made in **Figure 7A**, where we consider the phylogenetic histories of two closely related ecotypes. The phylogenetic diversity within an ecotype is expected to have a history of boom and bust cycles: one most-fit genetic variant emerges successfully from a periodic selection event; from that variant many descendant variants diverge from one another (a boom of diversity); then a new periodic selection event purges all variants within the population except that containing the adaptive mutation (the bust). After an ancestral ecotype splits into two ecotypes, the two ecotypic lineages each have their own private periodic selection events, where each purges the diversity only within its own ecotype. The recurrent purging of diversity within each ecotype results in each ecotype having relatively little sequence diversity (at any locus sampled), with a high level of sequence divergence between members of different ecotypes. Moreover, each ecotype is clearly a monophyletic group. Two ecotypes that have coexisted for a long time, as irreversibly separate entities, should be discernible as separate sequence clusters because there has been adequate time for neutral mutations to accumulate distinctly in each lineage (Palys et al. 1997).

This ecotype-based model of bacterial diversity predicts a nearly one-to-one correspondence between ecologically

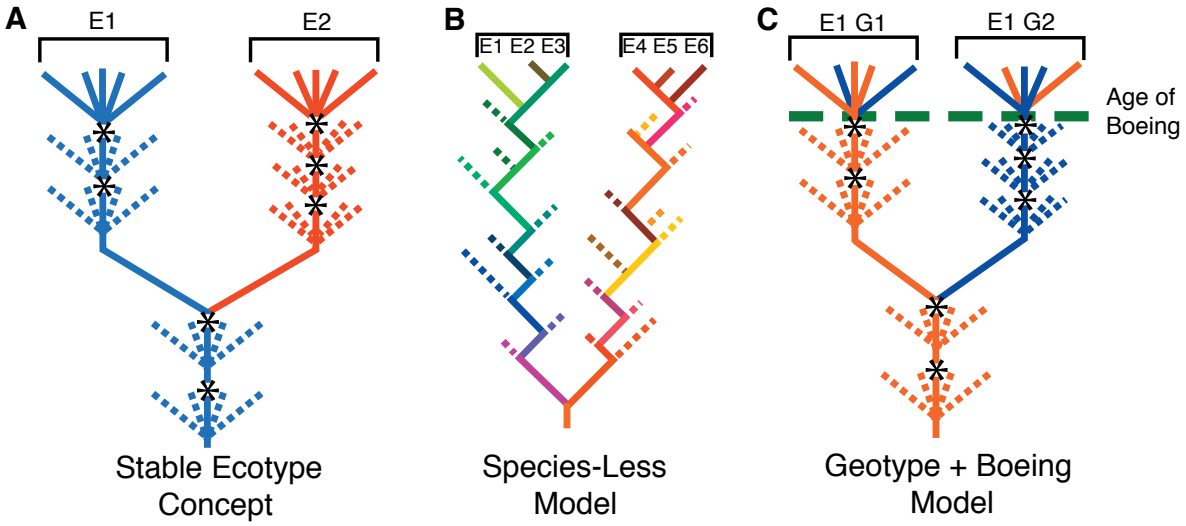


Figure 7. A. The phylogenetic history of two closely related ecotypes, under the Stable Ecotype Model. After each periodic selection event, indicated by an asterisk, only one variant from an ecotype survives. After periodic selection, the descendants of the surviving variant diverge (indicated by dashed lines), but with the next periodic selection event, again only one variant survives. Note that a sequence-based phylogeny of two ecotypes will indicate very limited sequence diversity within an ecotype, with much greater sequence divergence between members of different ecotypes. This model yields a 1-to-1 correspondence between ecotypes and sequence clusters. **B.** In the Species-less Model new ecotypes are invented and extinguished at a high rate. The diversity within each ecotype may be limited more by its short longevity (dashed lines indicate extinction) than by periodic selection. This model yields a many-to-1 relationship between ecotypes and sequence clusters. Each ecotype is depicted by a different color. **C.** In the Geotype-plus-Boeing Model geographic isolation in the past has resulted in multiple endemic sequence clusters within a single ecotype; then in the last few decades (above the dashed green line), jet travel has carried all the endemic clusters of the ecotype into every location. Eventually, periodic selection will collapse the worldwide diversity within the ecotype, but in this transitional period when air travel is still new, some ecotypes within a community may contain multiple sequence clusters. In panel C, there is only one ecotype; here the colors correspond to geographical location. In each panel, the different ecotypes are labeled as E1, E2, etc.; in panel C the different geotypes within Ecotype 1 are labeled E1 G1 and E1 G2.

distinct populations and sequence clusters provided that two assumptions are met: the ecologically distinct populations must be longstanding in their coexistence, and geographical isolation must not be the cause of sequence clustering.

3.3.2. Alternatives to the Stable Ecotype Model. Consider next the Species-Less Model, in which new ecotypes are constantly being invented (perhaps through horizontal transfer), while other ecotypes are going extinct (Boucher et al. 2001; Cohan 2004; Gogarten et al. 2002; Lawrence 2002; **Figure 7B**). Note that in this model, the sequence diversity within an ecotype may be constrained entirely by the ecotype's short longevity; in this case, there need not be a cohesive force (such as periodic selection) that can limit the diversity within the ecotype into the indefinite future (Godreuil et al. 2005). This species-less model of prokaryote diversity would yield a many-to-one correspondence between ecologically distinct populations and discernible sequence clusters.

In the Geotype Model, where migration between regions is very rare, geographically isolated populations of the same ecotype can evolve into distinct sequence clusters, a situation that applies frequently for non-dispersive animals and plants (Nelson and Platnick 1980; **Figure 7C**). This would yield a one-to-many relationship between ecotypes and endemic sequence clusters.

It is sometimes difficult to rule out the Geotype Model even when bacterial sequence clusters are sampled from the same geographic region. In a Geotype-plus-Boeing Model (**Figure 7C**), organisms of similar ecology (i.e., same ecotype) represented by sequence clusters from different regions (thus geographically diverged) may have migrated only recently into the same region (e.g., aided by airplanes); the various sequence clusters, now living in the same place, may not yet have endured a periodic selection event that would purge diversity throughout all clusters within the ecotype (Cohan 2004, 2005; Godreuil et al. 2005). In this case, even in a single community there would be a one-to-many relationship between ecotypes and sequence clusters.

In summary, we envision three different types of relations between ecotypes and sequence clusters within a single community: (i) Stable Ecotype Model—if ecotypes have the cohesive properties of species, owing to periodic selection, and are longstanding in their coexistence and not subject to geographic isolation, the relationship between ecotypes and sequence clusters within one community is expected to be nearly one-to-one; (ii) Species-less Model—if ecologically distinct populations are invented and extinguished at a high rate, with or without periodic selection there will be a many-to-one relationship between ecologically distinct populations and sequence clusters; and (iii) Geotype-plus-Boeing Model—if formerly geographically isolated populations of the same ecotype have only recently become sympatric (coexisting), perhaps with human transport, there will be a one-to-many relationship between ecologically distinct populations and sequence clusters. In our present research in Yellowstone, we aim to test whether the ecologically distinct *Synechococcus* populations of hot spring microbial mats have the cohesive properties of longstanding species.

3.3.3 Challenges of sequence-based discovery of ecotypes. Discovery of putative ecotypes as sequence clusters would seem straightforward enough. One would first identify sequence clusters and ask if each corresponds to an ecologically distinct population. The problem, however, is that within any group of closely related strains there is a hierarchy of clusters—subclusters within clusters, sub-subclusters within subclusters, and so on. It is not clear which of these clusters is expected to correspond to an ecotype.

We have developed an objective method for choosing the level of sequence cluster that is most likely to correspond to ecotypes (Cohan 2002d). The basic principle is that a population of huge size and with little recombination should have a phylogeny very similar to a star-shaped clade, with each contemporary strain sampled from the population descending directly from a single ancestral node (**Figure 8A, next page**). The most recent common ancestor of the ecotype represents the winner of the last periodic selection event, and owing to minimal genetic drift no two contemporary strains in a small random sample of the ecotype would be expected to be more closely related to one another than to any other stain. This is in

contrast to the case for animal species, in which frequent recombination eliminates the possibility of periodic selection, and small population sizes allow pairs of lineages to coalesce by genetic drift (**Figure 8B**). Nevertheless, even rare recombination within bacterial ecotypes can create additional nodes within an ecotype's phylogeny, especially if there is some recombination with strains outside the ecotype. It is a quantitative question how closely an ecotype's phylogeny should resemble a perfect star clade (i.e., with only one node), given the vagaries of recombination (Cohan 2002d).

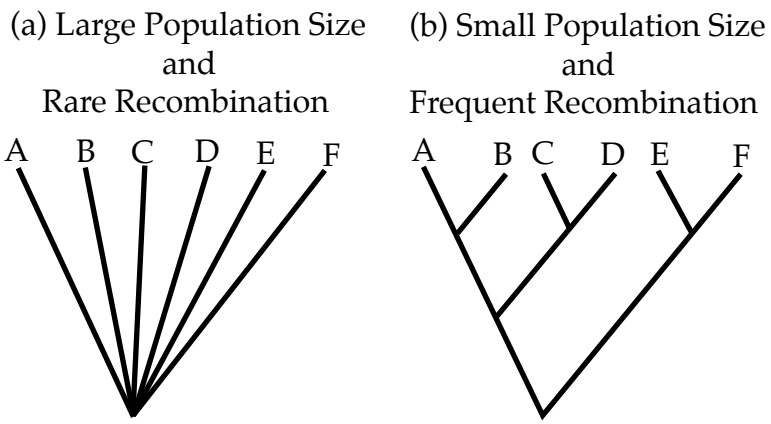


Figure 8. The phylogenetic signatures of populations whose diversity is controlled by periodic selection versus genetic drift (Cohan 2002a). Used with permission from *Ann Rev Microbiol.*

Our Star method is a computer simulation of sequence evolution within an ecotype that takes into account rates of mutation, recombination within and between ecotypes, periodic selection, and genetic drift—parameters that can be estimated from sequence data (Cohan 2002d). The aim is to find out the maximum number of nodes that can be expected within the phylogeny of a single ecotype given that the phylogeny is based on seven protein-coding genes. We have used protein encoding loci to gain high genetic resolution, and seven is the number of loci used in multilocus sequence typing (MLST) studies to group genetic variants into clonal complexes of related variants (Maiden et al. 1998; Spratt 1999). For example, we have previously found that the phylogeny of an ecotype within the rarely recombining *Staphylococcus aureus* is almost never expected to have more than one node; in the more frequently recombining *Neisseria meningitidis*, Star has shown that an ecotype's phylogeny frequently has one or two nodes, but almost never three or more (Cohan 2002d). This approach gives us a means for delimiting a putative ecotype. For example, if a group of strains within *N. meningitidis* has a phylogeny with three nodes, we should suspect that this group contains more than one ecotype. We have demonstrated that nearly all MLST clonal complexes within *S. aureus* and *N. meningitidis* have phylogenies consistent with that of a single ecotype (Cohan 2002d).

We are preparing to demarcate putative ecotypes within the hot spring cyanobacterial mat communities using the Star method. We will classify unique sequence variants (at protein-encoding loci) retrieved from the mat into putative ecotypes on the basis of their phylogenies, using the Star simulation as a guide to the number of nodes predicted within an ecotype's phylogeny. Next, we will attempt to determine whether these putative ecotypes are actually ecologically distinct entities. One clue will be the microhabitat of origin—temperature, quantity and quality of light and nutrients, oxygen concentration, pH, and perhaps other unknown variables that can change markedly over very small distances as well as over time within the mat, providing many possible spatiotemporal niches. As with 16S rRNA and ITS approaches above, the distinct distributions of putative ecotypes can indicate genetically based differences in optimal growth conditions. We will also compare putative ecotypes for their whole-genome patterns of gene expression, since changes in gene expression can bring about ecological divergence (Ferea et al. 1999). Finally, we will compare representatives of putative ecotypes to see if they possess different sets of genes in their genomes.

We will then attempt to determine whether the sequence clusters do indeed correspond to ecotypes, indicating that the ecologically distinct groups within the hot spring mats

are each a longstanding population, each with its own force of cohesion, and apparently irreversibly separate from all other ecotypes (supporting the Stable Ecotype Model). If we should find a many-to-one relationship between ecologically distinct populations and sequence clusters, we would conclude that there is a high turnover of species, with many newly invented species (i.e., Species-less Model). If we should find a one-to-many relationship, we would conclude that formerly geographically separated populations of the same ecotype have only recently been rejoined (Geotype-plus-Boeing Model). If each putative ecotype, as based on sequence clustering, is found to be ecologically homogeneous within itself but ecologically distinct from others, this would provide a structure for understanding the phenotypic and genomic diversity within the community. Any genomic or phenotypic diversity within an ecotype would likely represent merely the random neutral differences between ecologically interchangeable strains. The differences between ecotypes, on the other hand, deserve our fullest attention—these include the changes that underlie the community's ecological diversity, its structure, and its function.

3.4 Synthesis

Discrete distribution patterns of closely related 16S rRNA and ITS sequence variants along environmental gradients in hot spring cyanobacterial mats, combined with evidence that cultivated representatives of these variants are ecologically distinct, suggest that this variation arose through adaptive radiation. This patterning correlates well with the predictions of evolutionary ecology theory, which suggests that natural selection acts upon variation within a population to yield discrete, ecologically adapted populations. Interestingly, the word *ecotype* has been used independently by both microbial ecologists, who have observed diversity distribution patterns empirically, and by microbial evolutionary biologists, who use the word based on theoretical reasoning. Detection of ecotypes based on sequence variation in 16S rRNA genes and ITS regions may, however, be limited by the relatively conserved nature of these loci. We are developing high-resolution methods to study ecotype diversity in hot spring mats similar to the MLST methods used by microbial population geneticists. Periodic selection theory predicts that ecotypes that are long-lived and that are not influenced by geographic isolation can be detected as terminal evolutionary sequence clusters based on MLST locus sequence variation; each cluster that fits the phylogenetic predictions for a single ecotype can be delimited using an evolutionary simulation based on features of sequence data. The spatiotemporal distribution of variants within discrete sequence clusters along environmental gradients should distinguish among the Stable Ecotype Model and alternative models for how genetic diversity arises. If the Stable Ecotype theory explains the evolution of hot spring *Synechococcus*, we expect a 1:1 correspondence between cyanobacterial ecotypes and sequence clusters. However, if horizontal gene transfer causes rapid evolution of ecotypes (combined with their rapid extinction) we expect more than one ecotype per cluster. Geographic patterning of 16S rRNA and ITS variants suggests that physical isolation can also influence the divergence of prokaryotic populations. A significant geographic isolation effect may result in more than one cluster per ecotype within a single community. Thus, we are in a position to evaluate the extent to which competing theories can be applied to identify the fundamental ecological units of microbial communities.

ACKNOWLEDGEMENTS

We appreciate the support of the U.S. National Park Service and staff at Yellowstone National Park for granting permission for this research, as well as providing excellent support. These studies have been funded by grants from the National Science Foundation Ecology (BSR-9209677, BSR-9708136) and Frontiers in Integrative Biology Research (EF-0328698) Programs; the NASA Exobiology (NAG5-8824) and Astrobiology Institute (cooperative agreement CAN-97-01-OSS-004 with the NASA Ames Research Center (NCC 2-1073)); the Montana State University Thermal Biology Institute (NASA NAG5-8807); and Wesleyan University.

REFERENCES

Allewalt, J.A., M.M. Bateson, N.P. Revsbech, K. Slack, and D.M. Ward. Submitted. Temperature and light adaptations of *Synechococcus* isolates from the microbial mat community in Octopus Spring, Yellowstone National Park. *Appl Environ Microbiol.*

Atwood, K.C., L.K. Schneider, and F.J. Ryan. 1951. Periodic selection in *Escherichia coli. Proc Natl Acad Sci* 37:146-55.

Bauld, J., and T.D. Brock. 1973. Ecological studies of *Chloroflexis*, a gliding photosynthetic bacterium. *Arch Microbiol* 92:267-84.

Begon, M., J.L. Harper, and C.R. Townsend. 1990. *Ecology: Individuals, Populations, and Communities*, 2nd ed. Oxford, UK: Blackwell Science Ltd.

Boucher, Y., C.L. Nesbo, and W.F. Doolittle. 2001. Microbial genomes: dealing with diversity. *Curr Opin Microbiol* 4:285-9.

Brock, T.D. 1978. *Thermophilic microorganisms and life at high temperatures.* Berlin: Springer-Verlag.

Brock, T.D., and M.L. Brock. 1968. Relationship between environmental temperature and optimum temperature of bacteria along a hot spring thermal gradient. *J Appl Bacteriol* 31:54-8.

Claridge, M.F., H.W. Dawah, and M.R. Wilson. 1997. *Species: Units of biodiversity*. London: Chapman and Hall.

Cohan, F.M. 1994a. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Amer Naturalist* 143:965-86.

Cohan, F.M. 1994b. Genetic exchange and evolutionary divergence in prokaryotes. *Trends Ecol Evol* 9:175-80.

Cohan, F.M. 2001. Bacterial species and speciation. *Syst Biol* 50:513-24.

Cohan, F.M. 2002a. Clonal structure: an overview. In *Encyclopedia of Evolution*, ed. M. Pagel, 159-61. Oxford: Oxford University Press.

Cohan, F.M. 2002b. Population structure and clonality of bacteria. In *Encyclopedia of Evolution*, ed. M. Pagel, 161-3. Oxford: Oxford University Press.

Cohan, F.M. 2002c. Sexual isolation and speciation in bacteria. *Genetica* 116:359-70.

Cohan, F.M. 2002d. What are bacterial species? *Annu Rev Microbiol* 56:457-87.

Cohan, F.M. 2004. Concepts of bacterial biodiversity for the age of genomics. In *Microbial Genomes*, ed. C.M. Fraser, T.D. Read, and K.E. Nelson, 175-94. Totowa, New Jersey: Humana Press.

Cohan, F.M. 2005. Periodic selection and ecological diversity in bacteria. In *Selective Sweep*, ed. D. Nurminsky. Georgetown, Texas: Landes Bioscience.

de Queiroz, K. 1998. The general lineage concept of species, species criteria, and the process of speciation. In *Endless forms: species and speciation*, ed. D.J. Howard and S.H. Berlocher, 57-75. Oxford: Oxford University Press.

Feil, E.J., M.C.J. Maiden, M. Achtman, and B.G. Spratt. 1999. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Mol Biol Evol* 16:1496-1502.

Feil, E.J., J. Maynard Smith, M.C. Enright, and B.G. Spratt. 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* 154:1439-50.

Feil, E.J., J.E. Cooper, H. Grundmann, D.A. Robinson, M.C. Enright, T. Berendt, S.J. Peacock, J.M. Smith, M. Murphy, B. G. Spratt, C.E. Moore, and N.P. Day. 2003. How clonal is *Staphylococcus aureus*? *J Bacteriol* 185:3307-16.

Feldgarden, M., N. Byrd, and F.M. Cohan. 2003. Gradual evolution in bacteria: Evidence from *Bacillus* systematics. *Microbiol* 149:3565-73.

Fenchel, T. 2003 Biogeography for bacteria. *Science* 301:925-6.

Ferea, T.L., D. Botstein, P.O. Brown, and R.F. Rosenzweig. 1999. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc Natl Acad Sci* 96:9721-6.

Ferris, M.J., A.L. Ruff-Roberts, E.D. Kopczynski, M.M. Bateson, and D.M. Ward. 1996a. Enrichment culture and microscopy conceal diverse thermophilic *Synechococcus* populations in a single hot spring microbial mat habitat. *Appl Environ Microbiol* 62:1045-50.

Ferris, M.J., G. Muyzer, and D.M. Ward. 1996b. Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Appl Environ Microbiol* 62:340-6.

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. *Appl Environ Microbiol* 63:1375-81.

Ferris, M.J., M. Kühl, A. Wieland, and D.M. Ward. 2003. Different light-adapted ecotypes in a 68°C *Synechococcus* mat community revealed by analysis of 16S-23S intervening transcribed spacer variation. *Appl Environ Microbiol* 69:2893-8.

Futuyma, D.J. 1987. On the role of species in anagenesis. *Amer Naturalist* 130:465-73.

Godreuil, S., F.M. Cohan, H. Shah, and M. Tibayrenc. In press. Species concepts in microbes: Can we reconcile medicine and evolution? An E-debate. *Infect Genet Evol.*

Gogarten, J.P., W.F. Doolittle, and J.G. Lawrence. 2002. Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 19:2226-3

Hanada S, S. Takaichi, K. Matsuura, and K. Nakamura. 2002. *Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium which lacks chlorosomes. *Int J Syst Evol Microbiol* 52:187-93.

Koch, A.L. 1974. The pertinence of the periodic selection phenomenon to prokaryotic evolution. *Genetics* 77:127-42.

Kurtenbach, K., S. De Michelis, S. Etti, S.M. Schafer, H. Sewell, V. Brade, and P. Kraiczy. 2002. Host association of *Borrelia burgdorferi* sensu lato—the key role of host complement. *Trends Microbiol* 10:74-9.

Lawrence, JG. 2002. Gene transfer in bacteria: speciation without species? *Theor Pop Biol* 61:449-60.

Levin, B. 1981. Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. *Genetics* 99:1-23.

McGregor, K.F., B.G. Spratt, A. Kalia, A. Bennett, N. Bilek, B. Beall, and D.E. Bessen. 2004. Multilocus sequence typing of *Streptococcus pyogenes* representing most known *emm* types and distinctions among subpopulaion genetic structures. *J Bacteriol* 186:4285-94.

Maiden, M.C.J., J.A. Bygraves, E. Feil, G. Morelli, J.E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D.A. Caugant, I.M. Feavers, M. Achtman, and B.G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci* 95:3140-5.

Maynard Smith, J., N. Smith, M. O'Rourke, and B.G. Spratt. 1993. How clonal are bacteria? *Proc Natl Acad Sci* 90:4384-8.

Mayr, E. 1963. *Animal Species and Evolution*. Cambridge, MA: Belknap Press of Harvard University Press.

Mayr, E. 1982. *The growth of biological thought: diversity, evolution and inheritance.* Cambridge, MA: Harvard University Press.

Meglitsch, P. 1954. On the nature of species. *Syst Zool* 3:491-503.

Miller, S.R., and R.W. Castenholz. 2000. Evolution of thermotolerance in hot spring cyanobacteria of the genus *Synechococcus*. *Appl Environ Microbiol* 66:4222-9.

Nelson, G., and N.I. Platnick. 1980. A vicariance approach to historical biogeography. *Bioscience* 30:339-43.

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. *Appl Environ Microbiol* 68:4593-4603.

Palys, T., L.K. Nakamura, and F.M. Cohan. 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int J Syst Bacteriol* 47:1145-56.

Palys, T., E. Berger, I. Mitrica, L.K. Nakamura, and F.M. Cohan. 2000. Protein-coding genes as molecular markers for ecologically distinct populations: the case of two *Bacillus* species. *Int J Syst Evol Microbiol* 50:1021-8.

Papke, R.T., N.B. Ramsing, M.M. Bateson, and D.M. Ward. 2003. Geographical isolation and the evolution of hot spring cyanobacteria. *Environ Microbiol* 5:650-9.

Papke, R.T., and D.M. Ward. 2004. The importance of physical isolation to microbial diversification. *FEMS Microbiol Ecol* 48:293-303.

Peary, J.A., and R.W. Castenholz. 1964. Temperature strains of a thermophilic blue-green alga. *Nature* 202:720-1.

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. *Appl Environ Microbiol* 66:1038-49.

Revsbech, N.P., and D.M. Ward. 1984. Microelectrode studies of interstitial water chemistry and photosynthetic activity in a hot spring microbial mat. *Appl Environ Microbiol* 48:270-5.

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. *Appl Environ Microbiol* 68:1180-91.

Rosenzweig, M.L. 1995. *Species diversity in space and time*. Cambridge: Cambridge University Press.

Simpson, G.G. 1961. *Principles of Animal Taxonomy*. New York: Columbia University Press.

Spratt, B.G. 1999. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the Internet. *Curr Opin Microbiol* 2:312-16.

Templeton, A.R. 1989. The meaning of species and speciation: a genetic perspective. In *Speciation and Its Consequences*, 3-27. Sunderland, MA: Sinauer Associates Inc.

Turelli, M., N.H. Barton, and J.A. Coyne. 2001. Theory and speciation. *Trends Ecol Evol* 16:330-43.

Van Valen, L. 1976. Ecological species, multispecies, and oaks. *Taxon* 25:233-9.

Ward, D.M. 1998. A natural species concept for prokaryotes. *Curr Opin Microbiol* 1:271-7.

Ward, D.M., M.M. Bateson, M.J. Ferris, and S.C. Nold. 1998. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol Mol Biol Rev* 62:1353-70.

Ward, D.M., and R.W. Castenholz. 2000. Cyanobacteria in geothermal habitats. In *Ecology of Cyanobacteria*, ed. M. Potts and B. Whitton, 37-59. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ward, D.M., R.T. Papke, U. Nübel, and M.C. McKitrick. 2002. Natural history of microorganisms inhabiting hot spring microbial mat communities: clues to the origin of microbial diversity and implications for micro- and macro-biology. In *Biodiversity of Microbial Life: Foundation of Earth's Biosphere*, ed. J.T. Staley and A.-L. Reysenbach, 25-48. New York: John Wiley and Sons.

Wiley, E. 1978. The evolutionary species concept reconsidered. *Syst Zool* 27:17-26.