SHORT REVIEW

Genomics, environmental genomics and the issue of microbial species

DM Ward, FM Cohan, D Bhaya, JF Heidelberg, M Kühl, and A Grossman

A microbial species concept is crucial for interpreting the variation detected by genomics and environmental genomics among cultivated microorganisms and within natural microbial populations. Comparative genomic analyses of prokaryotic species as they are presently described and named have led to the provocative idea that prokaryotes may not form species as we think about them for plants and animals. There are good reasons to doubt whether presently recognized prokaryotic species are truly species. To achieve a better understanding of microbial species, we believe it is necessary to (i) re-evaluate traditional approaches in light of evolutionary and ecological theory, (ii) consider that different microbial species may have evolved in different ways and (iii) integrate genomic, metagenomic and genome-wide expression approaches with ecologic and evolutionary theory. Here, we outline how we are using genomic methods to (i) identify ecologically distinct populations (ecotypes) predicted by theory to be species-like fundamental units of microbial communities, and (ii) test their species-like character through in situ distribution and gene expression studies. By comparing metagenomic sequences obtained from well-studied hot spring cyanobacterial mats with genomic sequences of two cultivated cyanobacterial ecotypes, closely related to predominant native populations, we can conduct in situ population genetics studies that identify putative ecotypes and functional genes that determine the ecotypes' ecological distinctness. If individuals within microbial communities are found to be grouped into ecologically distinct species-like populations, knowing about such populations should guide us to a better understanding of how genomic variation is linked to community function. Heredity (2008) 100, 207–219; doi:10.1038/sj.hdy.6801011; published online 6 June 2007

Keywords: prokaryotic species; species concepts; environmental genomics; genomics; ecotype; microbial mat

Introduction

Understanding the nature of microbial species is an emerging issue in biology. While opinions differ widely on what constitutes a microbial species, they all have to take into account a universal pattern of diversity – that all organisms, including prokaryotes, fall into clusters of similarity in phenotype, sequence identity, genome content and ecology. One end of the discussion argues that there is no need to impose a theory on species demarcation. Species, as they have been traditionally demarcated based on phenotypic properties are viewed as acceptable, so long as they contain strains that are closely related phylogenetically. In this view, it is of no concern that the prokaryotic species presently recognized by systematics are known to be heterogeneous in their ecology (Rosselló-Mora and Amann, 2001; Rosselló-Mora, 2003). Others, however, view prokaryotic species as groups subject to specific dynamic forces and are trying to model species based on evolutionary forces suggested from empirical patterns of genetic variation. One such model is motivated primarily by the extensive evidence of horizontal gene transfer (HGT) and homologous recombination observed among close relatives, which some see as being in conflict with neo-Darwinian concepts (Rodríguez-Valera, 2004). This view suggests that species are groups of ecologically heterogeneous bacteria whose diversity is constrained by HGT and homologous recombination (Gogarten et al., 2002; Lawrence, 2002). Another model, is motivated by the in situ patterning of genetic variation, population genetic studies and evolutionary–ecology theory. This model suggests that the clusters among close relatives are ecologically distinct and irreversibly separate, and, as such are the fundamental units of diversity, much like species of animals and plants. We have hypothesized that there are such ecologically distinct, species-like groups of bacteria and we believe it is essential to identify and define these populations if we are to develop a predictive understanding of microbial community composition, structure and function (Ward and Cohan, 2005; Cohan, 2006; Ward, 2006; Ward et al., 2006). Here, our goal is to provide rationale for and introduce a theory-based approach we are taking to evaluate whether genomic variation in a natural microbial
community is organized into discrete phylogenetic and ecological clusters and whether such clusters exhibit properties expected of species.

Why should microbiologists care about species?

Plant and animal species are viewed as coherent clusters of individuals held together by cohesive forces, such as the ability to exchange genes (Templeton, 1989). There is a strong sense of the importance of species to communities: ‘the species...is 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’ (Mayr, 1982). Determining the complement of species and their abundances is the starting point for macroecological studies of community biodiversity, composition, structure, function and assembly (Begon et al., 1990; Hubbell, 2001). For instance, species rank abundance diagrams are used to define community structure or to assay changes in communities that result from dispersal, extinction, succession or anthropogenic change. Furthermore, to the extent that each species occupies a unique niche, these fundamental units serve unique roles and thus, link community composition and structure to community function. It has been suggested (Begon et al., 1990) that the goals of ecology are description, explanation, prediction and control. Description and explanation of species, viewed as populations of individuals that occupy unique niches and undergo unique dynamics (and interactions), is thus essential for predicting (i) how both transient perturbations and chronic environmental changes can modulate the structure and function of microbial communities and the interactions among community members, and (ii) how to control microbial communities for human-centric purposes.

Thoughts about microbial species

The ways that microbiologists view the species concept for prokaryotes have been reviewed recently (see papers in Claridge et al., 1997; Gevers et al., 2005; Cohan, 2006; Spratt et al., 2006; Ward, 2006; and see below). In this section, we provide an overview of ideas concerning microbial species upon which our approach is based.

Are named prokaryotic species true species?

Historically, prokaryotic species were named and recognized based on phenotypic (for example, morphological and physiological traits) clustering (Goodfellow et al., 1997; Rosselló-Mora and Amann, 2001). Microbiologists’ confidence that these ‘named’ bacterial species have biological significance has colored the development of microbial systematics and ecology. As molecular methods became available and were used for species demarcation, they were merely calibrated to yield previously established phenotypic clusters: a 60–70% cutoff for DNA–DNA hybridization, a 97–98% cutoff for similarity in 16S rRNA sequence, and most recently, a ∼94% average nucleotide identity (ANI) of genomes (Venter et al., 2004; Konstantinidis and Tiedje, 2005) were chosen as appropriate thresholds for grouping individu-
types, some >97% identical, are present in different regions of the mat (A/B-lineage genotypes in Figure 2a); all are distantly related to a readily cultivated *Synechococcus* (OS C1 isolate in Figure 2a). Clearly, a species concept based on morphological differences is meaningless for these organisms and cultivation biases have severely limited our ability to characterize the naturally abundant community members. If we were to apply a 2 or 3% 16S rRNA divergence cutoff to the classification of cyanobacteria in the mat, we would conclude that the variants that cluster into the *Synechococcus* A and B subclades represent strains within two species. However, each of the *Synechococcus* A/B 16S rRNA genotypes exhibits a unique distribution along the thermal gradient in the mat and isolates with these genotypes have been shown to be uniquely adapted to specific temperatures (Figure 2a) (Allewalt et al., 2006). Such patterning provoked us to consider concepts based on species being ecologically distinct populations (we will use the terms ecologically distinct populations, ecological species and ecotypes interchangeably, except where noted; Ward, 1998). Furthermore, using fast-evolving, high-resolution genetic markers, we (Palys et al., 2000; Ferris et al., 2003) and others (Zinser et al., 2006) have demonstrated that ecologically distinct populations may even have identical 16S rRNA sequences. For example, Ferris et al. (2003) had to use the internal transcribed spacer (ITS) that separates the 16S and 23S rRNA genes to genetically distinguish phenotypically distinct *Synechococcus* populations that inhabit different vertical zones in the mat (Ferris et al., 2003) (Figures 1 and 2b). Therefore, the ability to discern ecologically distinct populations within a community based on molecular methods depends on the degree of genetic resolution.

Investigations based on analysis of multiple genetic loci have demonstrated homologous recombination among closely related isolates from some microbial communities. For instance, Papke et al. (2004) and Whitaker et al. (2005) obtained evidence for high rates of homologous recombination in saltern *Halorubrum* and acid hot spring *Sulfolobus islandicus* isolates, respectively. Since high levels of homologous recombination have the potential to erode the distinction between suspected species populations (Hanage et al., 2005, 2006), it is important that we keep an open mind as to how different forces that generate and/or act upon variation are involved in speciation. Also, differing patterns of population genetics highlight the fact that we cannot expect all microbes to evolve in the same way.

Theoretical models of prokaryotic species and speciation

Sexual isolation is of paramount importance in isolating gene pools of sexual species (that is, Biological Species
As an alternative approach, microbiologists have often considered the species problem to be one of ‘where to draw the line’ that demarcates species (in terms of sequence divergence) as they evolve apart from one another (Ward, 2006). However, this is a simplistic approach since younger species will have diverged less than older, well-established species (Ward et al., 2006), different species are likely to speciate in different ways (Cohan, 2006) and relatively small changes in a genome (for example, the introduction of a single gene with a specific biological function, the elimination of gene function or even point mutations) can result in significant changes in functionality and fitness that result in a change of ecological niche (Rosenzweig et al., 1994; Rainey and Travisano, 1998). Therefore, we believe that it is more a matter of ‘how to draw the line’. Cohan (2006) emphasizes the importance of theoretical underpinnings, based on principles of evolution and ecology (see also Whitaker and Banfield, 2005), for understanding prokaryotic species, and has introduced a diversity of models (Table 1) that take into account the various ways in which diversity may arise within microbial populations (for example, via mutation, recombination and plasmid transfer) and how the environment may influence variation (for example, genetic drift, purifying selection, positive selection, geographic isolation and perhaps recombination. (Interestingly, based on evidence of high homologous recombination in populations they have studied, Whitaker and Banfield (2005) and Whitaker et al. (2005) suggest that recombination may sometimes also be a force acting upon prokaryotic variation by preventing divergence of new lineages, akin to sexual recombination in plants and animals.) (Figure 3). Different models may apply to speciation of different prokaryotes.

In our initial analysis of diversity in cyanobacterial populations of hot spring mats we are using the Stable Ecotype Model, since its predictions match the ecological patterning of molecular diversity that we have observed along environmental gradients (Figure 2a). In this model, (Concept) but, with the exception of highly recombining organisms, it is unlikely to be the cohesive force holding together asexual (or rarely sexual) species (Cohan, 1994). As an alternative approach, microbiologists have often considered the species problem to be one of ‘where to draw the line’ that demarcates species (in terms of sequence divergence) as they evolve apart from one another (Ward, 2006). However, this is a simplistic approach since younger species will have diverged less than older, well-established species (Ward et al., 2006), different species are likely to speciate in different ways (Cohan, 2006) and relatively small changes in a genome (for example, the introduction of a single gene with a specific biological function, the elimination of gene function or even point mutations) can result in significant changes in functionality and fitness that result in a change of ecological niche (Rosenzweig et al., 1994; Rainey and Travisano, 1998). Therefore, we believe that it is more a matter of ‘how to draw the line’. Cohan (2006) emphasizes the importance of theoretical underpinnings, based on principles of evolution and ecology (see also Whitaker and Banfield, 2005), for understanding prokaryotic species, and has introduced a diversity of models (Table 1) that take into account the various ways in which diversity may arise within microbial populations (for example, via mutation, recombination and plasmid transfer) and how the environment may influence variation (for example, genetic drift, purifying selection, positive selection, geographic isolation and perhaps recombination. (Interestingly, based on evidence of high homologous recombination in populations they have studied, Whitaker and Banfield (2005) and Whitaker et al. (2005) suggest that recombination may sometimes also be a force acting upon prokaryotic variation by preventing divergence of new lineages, akin to sexual recombination in plants and animals.) (Figure 3). Different models may apply to speciation of different prokaryotes. In our initial analysis of diversity in cyanobacterial populations of hot spring mats we are using the Stable Ecotype Model, since its predictions match the ecological patterning of molecular diversity that we have observed along environmental gradients (Figure 2a). In this model,
the cohesive force holding members of a species population together is provided by periodic selection (Figure 3). Diversity within a population arises through mutation and recombination, but recombination between populations is considered too rare to prevent adaptive divergence between ecologically distinct populations (Cohan, 1994). From time to time, the within-population diversity is purged by periodic selection events. In such a case, one new mutant or recombinant is fitter than the rest in the context of changing environmental conditions.

Table 1 Theoretical models of prokaryotic species and speciation that are not based on the assumption that named prokaryotic species are true species (Gevers et al., 2005; Godreuil et al., 2005; Ward and Cohan, 2005; Cohan, 2006)

| Model            | Description                                                                 | Predicted relationship between ecotypes and clusters | Cause of homogeneity within clusters | Example microbial group                      |
|------------------|-----------------------------------------------------------------------------|------------------------------------------------------|-------------------------------------|---------------------------------------------|
| Stable ecotype   | Periodic selection limits the diversity within an ecologically distinct population (ecotype) occupying a specific niche. Niche invasion mutations effect the evolutionary origin of an ecologically distinct variant that founds a new population. The ecotypes then accumulate their own mutations, and each population is purged of its diversity through its own private periodic selections. The result is two distinct phylogenetic clusters representing the ecologically distinct populations. | 1 ecotype : 1 cluster | Periodic selection | Mycobacterium tuberculosis (Smith et al., 2006); Bacillus simplex (Sikorski and Nevo, 2005) |
| Geotype+Boeing    | Same as the stable ecotype model, except that divergence occurs also due to the accumulation of neutral mutations in physically isolated populations. If these populations later come to inhabit the same community, there will be > 1 cluster per ecotype. | 1 ecotype : > 1 cluster | Dispersal within geographic regions, but not between regions | Yersinia pestis (Achtman et al., 2004) |
| Genetic drift    | Random fixation of allelic variants when population sizes are small, yielding clusters of close relatives within an ecotype. This is most likely for obligate pathogens and commensals in which host infections are founded by very few cells | 1 ecotype : > 1 cluster | Genetic drift | Buchnera (Herbeck et al., 2003) |
| Cohesive recombination | Multiple closely related ecotypes genetically recombine at a high rate resulting in a reduction in the distinctness of ecotypes. | Nondistinct clusters | Recombination between ecotypes | Helicobacter pylori (Falush et al., 2001) |
| Species-less     | Frequent invention and extinction of ecologically distinct populations and little periodic selection. This is most likely in a rapidly changing environment | Many ecotypes : 1 cluster | Young age of closely related ecotypes | |
| Recurrent niche invasion | Recurrent gain or loss of adaptations (for example, facile gain or loss of a plasmid) interferes with divergence of ecotype-specific lineages. | Many ecotypes : 1 cluster | Plasmid transfer | Rhizobium (Mutch and Young, 2004) Bacillus thuringiensis (Rasko et al., 2007) |
| Nano-niche Model | Subgroups within one ecotype each become adapted in nuanced ways to the subtleties of multiple microhabitats within a single system, but one adaptive mutant may still outcompete to extinct all the ecological diversity within the ecotype. The various ecologically distinct subgroups are not irreversibly separate, and do not have time to diverge into separate sequence clusters. | Many ecologically distinct populations : 1 cluster | Young age of subgroups; periodic selection within the whole ecotype | Vibrio splendidus (Thompson et al., 2005) |

*Ecotypes are defined as populations with distinct ecology; clusters are defined as monophyletic clades unifying variants predicted to be putative ecotypes.
that define the niche, and owing to the rarity of recombination, nearly the entire, intact genome of the original adaptive mutant sweeps through the population (with one known exception \((Helicobacter pylori; Falush et al., 2001)\), recombination rates are no greater than 10 times the mutation rate, which is insufficient to prevent periodic selection within an ecotype or disintegration of ecological divergence between ecotypes (Cohan, 2002b)).

The surviving, most-fit variant carries forward the general ecological character of the population occupying this niche and finds a new round of diversification that develops until another periodic selection event again eliminates within-population diversity. Thus, a species population evolves through a series of periodic selection events that recurrently purge the ephemeral diversity within the population. Permanent divergence occurs with the evolution of ecological novelty, either by accumulation of adaptive mutations or by sudden acquisition of genes via HGT, which founds a new species population. The founding variant and its descendants can escape the selection event that purges diversity in the parental population, and this new population can then undergo its own private periodic selections. In the Stable Ecotype Model, periodic selection is assumed to occur many times during the lifetime of an ecotype. Thus, the outcome of repeated private periodic selections in each of the two ecotype lineages is the evolution of discrete phylogenetic clusters that represent ecologically unique species. We refer to these clusters as ecotypes, but they could also be considered species using an ecological species concept (Ward, 2006) (ecotypes may be defined generally as ecologically distinct populations, or more specifically as ecologically distinct populations that are subject to their own private periodic selection events. It is this latter definition that allows us to identify ecotypes as clusters based on DNA sequence identity (see below)). In sum, the Stable Ecotype Model predicts a fundamental evolutionary discontinuity – closely related individuals cluster into species-like ecotypes that are acted upon uniquely by natural selection.

It is important, however, to take into account other models of microbial speciation. For example, an ecotype of small effective population size, dominated by genetic drift (Genetic Drift Model), may contain multiple sequence clusters. Ruling out genetic drift as a major force acting upon diversity seems reasonable in the hot spring mats since \(Synechococcus\) population sizes are enormous \((10^{10} cells/ml mat)\). We have also considered how the expected 1:1 relationship between phylogenetic clusters and ecologically distinct populations predicted by the Stable Ecotype Model might be altered as a consequence of geographic isolation (Geographic + Boeing model predicts > 1 cluster per ecotype) or very rapid HGT (for example, the Species-less Model predicts > 1 ecotype per cluster) (Ward and Cohan, 2005) (Table 1). We reiterate that the evidence in hand suggests that different microbes evolve in different ways. The Stable Ecotype Model is just a starting point and our analysis of genomic variation in \(Synechococcus\) populations will tell us the extent to which this and/or other models of species and speciation apply to the natural microbial populations that we are studying.

**Genomics and the species issue**

As implied in our previous discussion, comparative genomic analysis of strains of named prokaryotic species cannot lead to an accurate view of variation within a species if the strains compared are not really from a single true species population. As depicted in Figure 4, comparing different strains within a named species may be more like comparing strains belonging to different species within a genus. If, alternatively, we can determine which of the individual strains group into true species populations, we should be able to use genomic analysis to investigate variation among strains believed to be functionally interchangeable (that is, from the same ecotype) or among strains believed to be ecologically distinct (that is, from different ecotypes). Such analyses should help us to understand whether genomic differences (at the nucleotide or gene level) are ecologically
and evolutionarily neutral or instead define adaptive traits that distinguish different ecotypes. Recently, an interesting attempt was made to use comparative genomics to determine the ANI separating populations with ecological distinctions (Konstantinidis and Tiedje, 2005). It was suggested that such a molecular standard ‘could be as stringent as including only strains that show >99% ANI or are less identical…but share an overlapping ecological niche’. However, a single cutoff is not likely to be realistic since different prokaryotic species would be expected to evolve in different ways, and any species population evolves from a nascent stage (hardly any neutral differences) to more mature stages (increasingly greater neutral differences as new populations diverge from their parental populations).

Environmental genomics and the species issue
Environmental genomic analysis represents a major advance in our ability to discover the enormous diversity within microbial communities and, in some cases, to associate this diversity with ecological significance. Recent reviews and opinion papers, many of which we cite below, have been written about these methods, their promise for providing a deeper understanding of microbial communities and their pitfalls. The issue of microbial species is frequently raised in metagenomic papers because, in all studies, micro-heterogeneity, defined as sequence diversity among very closely related variants, has been observed (DeLong, 2005). Different authors express different opinions on the issue of microbial species, but most consider it important. As DeLong (2004) has said, ‘a critical parameter element, ‘microbial species’, is still looking for a concrete definition’ and as Newman and Banfield (2002) said ‘as we begin to sample the genomes of natural populations, we must confront the question of species- and subspecies-level genome diversity’.

The ‘gold standards’ for sequence divergence (discussed above) have been used in some metagenomic studies to describe the diversity and genetic character of species in microbial communities. For instance, Venter et al. (2004) used 97 and 99% rRNA sequence similarity cutoffs to estimate species diversity in their Sargasso Sea plankton metagenomic study; clearly acknowledging that ‘though sequence divergence does not universally correlate with the biological notion of ‘species’…sequence similarity within the rRNA genes is the accepted standard in studies of uncultured microbes’. A cutoff of >94% ANI for assembled and unassembled sequences, was used to define ‘genomic’ species since it corresponded to ~97% rRNA sequence identity; this criterion was used to conservatively demarcate species for species richness calculations. Similarly, Schleper et al. (1998) employed ‘standard criteria (for example, rRNA and genomic DNA similarity)’ to interpret clonal variants of an archean retrieved from a marine sponge that exhibit 99.2–99.3% rRNA and ~87% overall DNA sequence identity as ‘strains of a single species’. In their study of ‘Haloquadratum walsbyi’ inhabiting a salt crystallizer pond, Legault et al. (2006) also applied a 94% ANI to identify what they ‘considered as bona fide ‘H. walsbyi’ (belonging to lineages closely related to the sequenced strain). Although the title of their paper implies that this is a study of a single species, the use of quotation marks and reference to closely related lineages appears to reflect uncertainty about whether this named archaeal species is truly a single species.

Others have not limited their interpretation of metagenome data to the application of the standard molecular cutoffs. For instance, in their study of acid mine drainage biofilms, Tyson et al. (2004) assembled metagenome clones based on overlapping of nearly identical sequences and binned them based on G+C content and depth of coverage into a ‘composite genome’ called Ferroplasma (type II). (Individual clone sequences were assembled into contigs based on nearly-identical sequence overlaps, then scaffolds were built using individual clones whose ends overlap with different individual contigs; the entire ‘composite genome’ is based on binning scaffolds that show similar G+C content and depth of coverage.) They considered this composite genome to represent a new species distinct from that represented by an existing Ferroplasma isolate (type I) that was only 1% divergent at the 16S rRNA locus. Nucleotide polymorphisms occurring within the composite genome at a frequency of 2.2%, were considered to have come from differences among strains of the Ferroplasma II species. The Ferroplasma II strains they demarcated showed evidence of homologous recombi-
nation, which was not as evident between Ferroplasma I and II. Based on the expectation that homologous recombination is less frequent among more divergent populations (for example, species more divergent than strains), the authors suggested that 'recombination and assembly may provide useful genome-based criteria to separate species from strains in cases where one or both organisms are uncultivated'. (The authors make it clear that it is really the rate of recombination, not merely recombination.) In the same study, a composite genome of a Leptospirillum population (group II) that exhibited only 0.08% nucleotide polymorphism was considered to be representative of a single strain. Rather than being bound by the ‘gold standard’ molecular cutoffs, Tyson et al. (2004) apparently used the composite genomes to demarcate species (that is, species were demarcated by assemblies of individual sequences that were binned on the basis of their %G + C content falling within a range of ~10–15% and depth of coverage) and variants within these composite genomes were considered strains of this species. Although, this approach, like the use of molecular cutoffs, allows for theory-independent prediction of species based on metagenomic data, without further experimental work it remains unknown whether the populations predicted as species in this way correspond to populations with the expected properties of species (see above and below). Nevertheless, these authors were clearly thinking about the influence of evolutionary and ecological processes as they interpreted the patterns revealed by their data (Banfield et al., 2005; Whitaker and Banfield, 2005). For instance, they noted evidence consistent with a recent reduction in diversity in the Leptospirillum group II population, possibly due to a selective sweep (that is, period-specific selection event) or a founder effect. Whitaker and Banfield (2005) also pointed out that different population genetics patterns for Leptospirillum and Ferroplasma are consistent with the notion that different microbial lineages speciate in different ways and with differing frequencies (a founder effect is a type of population bottleneck, wherein population diversity becomes very low because of recent colonization of a new environment by a small number of individuals. Population bottlenecks might also have other causes (for example, near extinction events)).

The major issue that concerns us is whether, and if so how and why, closely related individuals cluster into species-like ecotype populations. Since high-resolution genomic and metagenomic approaches are needed to address this concern we have considered how such methods might best be used to resolve the issue of determining what microbial species are. Shotgun sequencing of DNA isolated from environmental samples represents an attempt to obtain sequence information for a large proportion of genes in a community, providing what has been described as a ‘parts list’ for a microbial community (DeLong, 2005). Recent gene-centric studies did not attempt to assemble the parts, but rather tried to identify functional ‘environmental gene tags’, yielding a coarse snapshot of community function that can be useful for determining general structural/metabolic/regulatory features of microbial communities (Tringe et al., 2005; DeLong et al., 2006). For understanding the genomic variation among individuals, however, the parts need to be assembled. Assembly of metagenomic sequences into contigs, scaffolds and their binning to form ‘composite genomes’ is complicated by many factors, including species diversity and evenness (DeLong, 2005; Schloss and Handelsman, 2005). (Evenness is the degree to which different species within a community occur in equal abundance.) All metagenomic assemblies to date include microheterogeneity and are really comprised of regions from a large number of individual genomes that are present in the environmental sample (that is, assembly and subsequent binning leads to composite or ‘virtual’ genomes). Hence, we cannot use metagenomic assemblies to study genomes of individuals, and, without further investigation we cannot assess whether the variation within assemblies corresponds to that found within one or more than one true species. Approaches using fosmid and bacterial artificial chromosomes (BACs) do, however, sample large segments of genomes of individuals, albeit small in comparison to the entire genome, and can link clusters of genes with diverse sequences to specific variants in the population. There are a number of elegant examples that demonstrate how this approach has led to the discovery of metabolic processes in groups of bacteria that were not known to conduct such metabolisms (for example, Béjà et al., 2000, 2001, 2002; Bryant et al., 2007). Cloning of large genomic regions has also been used to assay sequence microheterogeneity among individuals in marine crenarchaeal populations (Schleper et al., 1998; Béjà et al., 2002a, b) and ‘H. quadratricum’ in a salt crystallizer (Legault et al., 2006).

Our integrated theory-based approach

As a result of our experimental results and theoretical considerations, we envision ecotypes to be fundamental species-like units that occupy unique niches within microbial communities. As such, ecotypes should regulate community function through unique patterns of population distribution, and by the dynamics of gene expression and metabolic activity within the context of their environment. Based on predictions of the Stable Ecotype Model of speciation, we pose two general questions:

1. Is genomic variation within a natural microbial community organized into discrete phylogenetic and ecological clusters, as expected of ecotypes?

2. Do these clusters exhibit properties expected of ecotypes (discrete ecological distributions, discrete functions based on gene content and/or sequence adaptations, discrete patterns of gene expression)?

To address these questions, we are using genomic and population genetic methods to investigate the well-characterized hot spring microbial mats of Octopus Spring (Ward et al., 1998) and Mushroom Spring (Ward et al., 2006) (Figure 1). The mat communities are ideal for such analyses: (i) they have an uneven community structure skewed toward large, predominant type-A/B Synechococcus populations, (ii) genetically and ecologically relevant isolates, both axenic and nonaxenic, of the type-A/B populations are available (Allewalt et al., 2006; Kilian et al., 2007), (iii) the mats have well-defined temperature, light and chemical gradients that can be measured at the microscale level using microsensors that can also quantify photosynthesis and other microbial activities in situ (Ward et al., 2006) (Figure 1), (iv) these gradients can be experimentally subsampled (Ramsing...
et al., 2000; Ferris et al., 2003), (v) the mats have very high biomass, are readily accessed, and are protected within Yellowstone National Park, (vi) there is background information on the predominance and distribution of the specific 16S rRNA and ITS genotypes in the mat and (vii) previous studies have explored cyanobacterial physiology over the diel cycle (for example, van der Meer et al., 2005, 2007). This strong foundation allows for the development of rational hypotheses related to studies of acclimation and adaptation within and between putative ecotype populations.

Genomic and metagenomic databases
Because unraveling the prokaryotic species issue requires high-resolution approaches, we have used genomic and metagenomic analyses to probe the diversity for all genes of the dominant phototrophic members of the community. We have obtained the complete genome sequences for two Synechococcus isolates (Bhaya et al., 2007, www.tigr.org), one A and one B' genotype. These isolates were obtained from Octopus Spring and differ in their adaptation to temperature in correspondence with the distribution of these ecotypes along the thermal gradient (Allewalt et al., 2006) (Figure 2a). We have also constructed metagenomic libraries from the ñ1 mm thick top green layer of the mats containing all of the Synechococcus cells, as well as other microorganisms present in these samples (Bhaya et al., 2007). Separate libraries were constructed from samples collected from two sites defined by average temperatures of 60 and 65 °C at Octopus and Mushroom Springs, where Synechococcus 16S rRNA genotypes A and B' or just genotype A occur, respectively. By using the sequenced genomes as anchors onto which metagenome sequences are positioned, we have initiated an analysis of Synechococcus A-like and B'-like variants inhabiting the mats (Bhaya et al., 2007). This represents a powerful approach for identifying variants in the population. For instance, we can detect potential recombination events between closely related populations as well as discover genes that are present in one sub-population but not in another. Furthermore, through in situ gene expression analysis (see below) we can determine whether or not these genes are transcribed (Steunou et al., 2006). The anchor genomes also allow us to identify sequences from Synechococcus A-like and B'-like variants, thus facilitating theory-based analysis of how these populations may be subdivided into putative ecotypes (Ward et al., unpublished). It may eventually be possible to link functional variants with ecotypes identified through population genetics studies.

Theory-based prediction of putative ecotypes
A central and distinguishing theme of the work that we have initiated involves identification of putative ecotype populations using genomic/metagenomic information coupled to theory-based approaches. We have developed and implemented an ‘ecotype simulation’ analysis method that is based on simulating the evolutionary history of a phylogenetically defined group (clade) within a community as described by variation at individual genetic loci (Cohan, 2006; Cohan and Perry, 2007). Rather than applying an arbitrary cutoff, the phylogeny of the organisms under study is used to estimate parameters responsible for the evolution of this particular clade (periodic selection rate, ecotype formation rate, number of ecotypes). By identifying the smallest clusters consistent with single ecotypes, we can predict how individual variants are grouped into ecotypes. We have validated the ecotype simulation by analyzing data on the distribution of ITS variants present in the 68 °C region of the Mushroom Spring mat (Figure 2b) and have found that the model predicts ecotypes that correspond with sequence clusters that were previously shown to be associated with surface and subsurface layers of the mat (Ward et al., 2006). Cohan (2006) and Cohan and Perry (2007) provide other examples of correspondence between ecologically distinct sequence clusters and putative ecotypes predicted by ecotype simulation analysis. We have also begun to use more rapidly evolving protein-encoding genes for the ecotype simulation analysis, which offer greater molecular resolution.

We are also developing cultivation-independent MLST methods in which we have constructed large-insert metagenomic libraries from mat DNA using BAC clones. This permits us to sample variation over multiple loci contained within >100 kb segments of the genomes of individuals within native Synechococcus populations. Our strategy is to select multiple loci that have maximally diverged among A/B-type Synechococcus populations to increase molecular resolution. Analysis of variation at these multiple loci will also buffer against the potential challenge of frequent homologous recombination, whereby a single-locus phylogeny may fail to resolve species populations (Hanage et al., 2005, 2006).

We suggest that ecotypes be hypothesized on the basis of falling into distinct sequence clusters that are predicted on theoretical grounds to have had a history of coexistence, and that these putative ecotypes be confirmed by demonstrating their ecological distinctness (Cohan, 2006). Staley (2006) has suggested a phylogenetic species concept for prokaryotes that defines species, however they may have formed (for example, through adaptation, geographic isolation), as the smallest ‘irreducible’ phylogenetic clusters. Meeting this criterion requires that we examine putative ecotypes using genetic markers with increasing resolving power to identify the point at which increasing molecular resolution no longer divides a putative ecotype cluster into more than a single ecotype.

Testing whether putative ecotypes exhibit species-like properties
Since, as ecotype populations diverge, all but the most conserved genes should accumulate neutral genetic differences, it should be possible to define ecotype-specific allele sets for most Synechococcus genes, including highly expressed functional genes (Figure 3). (The use of functional genes as neutral markers may not seem intuitively obvious, since it is tempting to think of functional genes as being the genes under selection. However, we are focusing on neutral differences, not adaptive differences. By examining molecular details (for example, $K_s$/$K_a$ ratio), it is possible to identify and thus avoid genes are under strong positive selection.) These sets of allelic variants can be used to evaluate whether putative ecotypes exhibit the ecological distinctness expected of species-like populations. This is essentially
the same kind of approach we used to determine whether 16S rRNA and ITS variants have unique ecological distributions relative to the horizontal flow (for example, temperature, nutrients) and vertical gradients (for example, light and chemical parameters that covary with light, and nutrients) in the mats. As an example of our approach, recall that the phenotypically distinct *Synechococcus* populations that inhabit the uppermost and deeper portions of the 68°C Mushroom Spring mat photic zone (Figure 1) show differences in ITS sequence (Figure 2b). Although similar phenotypically distinct *Synechococcus* populations are found at different depths in the 65°C mat photic zone (Ward et al., 2006), these populations do not exhibit variation in ITS sequence. We might hypothesize that all of these *Synechococcus* cells belong to a single ecotype, with cells in different microenvironments differently acclimated (for example, to different light regimes). Alternatively, we might hypothesize that there are two (or more) even more closely related, yet distinctly adapted ecotypes. If there is a single ecotype, we would expect ecotype simulation and/or MLST analysis of yet more high-resolution loci to predict just one putative ecotype cluster and allelic variants that define this cluster should be found at both depths. If there are multiple ecotypes, we would expect >1 predicted putative ecotype cluster and the allelic variants that are unique to each cluster should be uniquely distributed to distinct depth intervals. It is conceivable that multiple ecotypes could have the same spatial distribution yet exhibit different temporal activity patterns, or that it might be impossible to sample on the spatial scale needed to resolve adjacent very small in situ spatial niches. To circumvent these challenges, and to assess the possibility of acclimation, we will examine spatiotemporal patterning of expression of ecotype-specific alleles. This, of course, can only be performed if it is possible to measure in situ gene expression.

**In situ gene expression**

We have made considerable progress with a variety of approaches to assess in situ gene expression. We have targeted the analyses of specific transcripts using reverse transcriptase PCR amplification and quantitative PCR. These transcripts encode proteins that perform important physiological processes in *Synechococcus*, including photosynthesis, respiration, fermentation and nitrogen fixation (genes encoding Nif proteins were discovered in the initial *Synechococcus* genomic analyses; Steunou et al., 2006). Strong shifts in the abundances of these transcripts were observed during a natural light to dark transition. Transcripts encoding proteins involved in photosynthesis and respiration are highest during the day while those encoding proteins associated with fermentation and nitrogen fixation accumulate to high levels in the evening (Figure 5). By combining gene expression studies with microsensor analysis of environmental parameters (Figure 5), we can begin to identify controls on gene expression of metabolic processes in the mat. While the *nif* genes are likely controlled primarily by oxic conditions, the photosynthesis genes are light responsive (even under anoxic conditions), and the respiratory genes may be under circadian control. Additionally, analyses of axenic *Synechococcus* OS-B' cultures under defined conditions are demonstrating specific light effects on growth, pigmentation and gene expression (Kilian et al., 2007) that will inform experiments to be performed in situ. Finally, we have developed a microarray that contains oligonucleotide primers (70 mers) representing genes on the *Synechococcus A* and *B* genomes; we will thus be able to simultaneously evaluate expression (for example, both A and B' genotypes co-occur in the 60°C mat) of most genes in these organisms, and potentially other variants present in the mat samples. This will constitute a first step toward ecotype-specific gene expression and will guide us to other highly expressed, high-resolution genes that may be useful for studies of yet more closely related *Synechococcus A/B* ecotypes defined by population genetics studies.

**The value of an evolutionary and ecological view of microbial species**

The biological implications of viewing microbial species using evolutionary and ecological principles coupled with genomic information are enormous. To the extent
that species are unique ecologically adapted populations, they may constitute the basic building blocks from which guilds and communities are assembled and they may be the populations whose dynamics vary in response to environmental changes (guilds are functional units of microbial communities comprised of many species that all do the same function in somewhat different ways (for example, many plant species conduct photosynthesis within a forest community)). Ecological species could represent populations with unique gene assemblages, and thus play pivotal roles in regulating community function in space and time, according to how they are spatially distributed, how their gene expression varies temporally and how environmental changes alter their abundances. They may also be the populations that co-evolve in response to biotic niche determinants. Since microbial communities play major roles in all ecosystems, gaining knowledge of their composition, structure and function has widespread predictive importance (Staley et al., 1997). An evolutionary and ecologically grounded view of species could also guide microbiologists toward a more orderly view of genome evolution—gene order and content among closely related microorganisms may not be as chaotic as it seems for named species, but rather may be ordered by the evolution of distinct ecotypes. Genomes from different ecotypes may differ in the sets of genes they have obtained through HGT: some of these genes would likely confer ecological uniqueness and may define an ecotype, while others may represent neutral changes that are not ecologically meaningful; genomes from the same ecotype should differ only in the latter set of genes (Cohan, 2006). In essence, like macrobiologists (for example, http://www.egad.ksu.edu/ and http://mimulusevolution.org/index.php), microbiologists must determine the pattern of suspected species populations before they can make sense of the underlying differences that matter in terms of ecological function. A major challenge will be to develop ways to identify and resolve differences in ecological function, whether they be due to adaptations to specific physical/chemical/biological niche determinants incorporated into gene/protein sequences (Miller, 2003; Bielawski et al., 2004) or to gain or loss of functionally critical genes (Ochman and Moran, 2001). Novel properties (for example, HGT, rapid doubling time) may influence the frequency and manner in which variation is generated in prokaryotes relative to plants and animals, but the way in which selection acts upon this variation may not differ – that is, there may be unifying principles of evolution, ecology, physiology and molecular biology across scales of size and the complexity of organisms.

Acknowledgements

Our research is funded by the National Science Foundation Frontiers in Integrative Biology Program (FIBR) award EF-0328698. MK also acknowledges additional support from the Danish Natural Science Research Council. We thank the many FIBR students, postdocs, research associates, technicians and colleagues who have worked hard to implement the study we have outlined here, as well as participants in our annual FIBR workshops (http://landresources.montana.edu/FIBR/), with whom we have had stimulating discussions about the ideas we express: Mary Bateson, Melanie Melendrez, Natsumo Hamamura, Chris Klatt, Eric Becraft, John Peters and Eric Brecht (MSU), Alex Koeppl, Danny Krizanc, and Liz Perry (Wesleyan), Anne Soisig Steunou, Oliver Kilian, Melissa Adams and Natalia Khuri (Carnegie), Sheila Jensen and Anni Glud (MBL, Copenhagen) and William Nelson (TIGR). We also thank the US National Park Service and personnel from Yellowstone National Park for their permission to conduct this work and their helpful assistance.

References

Achtman M, Morelli G, Zhu P, Wirth T, Diehl I, Kusecek B et al. (2004). Microevolution and history of the plague bacillus, Yersinia pestis. Proc Natl Acad Sci USA 101: 17837–17842.
Alleva`li JF, Bateson MM, Revsbech NP, Slack K, Ward DM (2006). Effect of temperature and light on growth and photosynthesis of Synechococcus isolates typical of those predominating in the Octopus Spring microbial mat community. Appl Environ Microbiol 72: 544–550.
Banfield JF, Tyson GW, Allen EE, Whitaker RJ (2005). The search for a molecular-level understanding of the processes that underpin the Earth’s biogeochemical cycles. Revs Mineral Geochem 29: 1–7.
Begon M, Harper JL, Townsend CR (1990). Ecology: Individuals, Populations, Communities, 2nd edn. Blackwell Science: Oxford.
Béja O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP et al. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science 289: 1902–1906.
Béja O, Spudich EN, Spudich JL, Leclerc M, DeLong EF (2001). Prokaryothorhodopsin phototrophy in the ocean. Nature 411: 786–789.
Béja O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, Hamada T et al. (2002a). Unsuspected diversity among marine aerobic anoxygenic phototrophs. Nature 415: 630–633.
Béja O, Koonin EV, Aravind L, Taylor LT, Seitz H, Stein JL et al. (2002b). Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. Appl Environ Microbiol 68: 335–345.
Bhaya D, Grossman AR, Steunou A-S, Khuri N, Cohan FM, Hamamura N et al. (2007). Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME J (in press).
Bielawski JP, Dunn KA, Sabebi G, Béja O (2004). Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. Proc Natl Acad Sci USA 101: 14824–14829.
Boucher Y, Nesbo CL, Doolittle WF (2001). Microbial genomes: dealing with diversity. Curr Opin Microbiol 4: 285–289.
Bryant DA, Garcia Costas AM, Maresco JA, Gomez A, Chew M, Klatt CG et al. (2007). Candidatus Chloroacidobacterium thermophylum: an aerobic phototrophic acidobacterium. Science (in press).
Claridge MF, Dawah HW, Wilson MR (eds) (1997). Species: Units of Biodiversity. Chapman and Hall: London.
Cohan FM (1994). The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. Am. Naturalist 143: 965–986.
Cohan FM (2002a). What are bacterial species? Ann Rev Microbiol 56: 457–487.
Cohan FM (2002b). Population structure and clonality of bacteria. In: M Pagel (ed). Encyclopedia of Evolution, vol. 1. Oxford University Press: New York. pp 161–163.
Cohan FM (2006). Toward a conceptual and operation union of bacterial systematics, ecology, and evolution. Phil Trans R Soc London B 361: 1985–1996.
Cohan FM, Perry E (2007). A systematics for discovering the fundamental units of bacterial diversity. Curr Biol 17: R373–R386.
Heredity

218

Godreuil S, Cohan FM, Shah H, Tibayrenc M (2005). Which

Giovannoni SJ, Stingl U (2005). Molecular diversity and ecology

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ

Lawrence JG (2002). Gene transfer in bacteria: speciation

Legault BA, Lopez-Lopez A, Alba-Casado JC, Doolittle WF, 

Kilian O, Steunou A-S, Fazeli F, Bailey S, Bhaya D, Grossman

The Unified Neutral Theory of Biodiversity and

Hanage WP, Fraser C, Spratt G (2006). Sequences, sequence

Falush D, Kraft C, Taylor-AG, Wieland A, Ward DM (2003). Different light-

Maiden MCJ, Byggraves JA, Feil E, Morelli G, Russell JE, Urwin

Dykhuizen DE (1998). Santa Rosalia revisited: why are there so

DeLong EF (2005). Microbial community genomics in the ocean.

Moran NA (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292: 1096–1098.

Palys T, Berger E, Mitryca I, Nakamura LK, Cohan FM (2000).

Protein-coding genes as molecular markers for ecologically distinct populations: the case of two Bacillus species. Int J Syst Evol Microbiol 50: 1021–1028.

Papke RT, Koenig JE, Rodriguez-Valera F, Doolittle WF (2004). Frequent recombination in a saltern population of Halorubrum. Science 306: 1928–1929.

Rainey PB, Fravisano M (1998). Adaptive radiation in a heterogeneous environment. Nature 394: 69–72.

Ramsing NB, Ferris MJ, Ward DM (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: 1038–1049.

Rasko DA, Rosovitz MJ, Okstad OA, Fouts DE, Jiang L, Cer RZ et al. (2007). Complete sequence analysis of novel plasmids from emetic and periodontal Bacillus cereus isolates reveals a common evolutionary history among the B. cereus-group plasmids, including Bacillus anthracis pXO1. J Bacteriol 189: 52–64.

Rodriguez-Valera F (2004). Environmental genomics, the big picture? FEMS Microbiol Lett 231: 153–158.

Rosenzweig RF, Sharp RR, Treves DS, Adams J (1994). Microbial evolution in a simple unstructured environment: genetic differentiation in Escherichia coli. Genetics 137: 903–917.

Rosselló-Mora R (2003). Opinion: the species problem, can we achieve a universal concept? Syst Appl Microbiol 26: 323–326.

Rosselló-Mora R, Amann R (2001). The species concept for prokaryotes. FEMS Microbiol Rev 25: 39–67.

Schleper CA, DeLong EF, Preston CM, Feldman RA, Wu K-Y, Miller SR (2003). Evidence for the adaptive evolution of the Bacillus anthracis pXO1. J Bacteriol 185: 6926–6936.

Schloss PD, Handelsman J (2005). Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. Genome Biol 6: 229 (doi:10.1186/gb-2005-6-6-229).

Sikorski J, Nevo E (2005). Adaptation and incipient sympatric speciation of Bacillus simplex under microclimatic contrast at ‘Evolution Canyons’ I and II, Israel. Proc Natl Acad Sci USA 102: 15924–15929.

Smith NH, Kremer K, Inwald J, Dale J, Driscoll JR, Gordon SV et al. (2006). Ecotypes of the Mycobacterium tuberculosis complex. J Theor Biol 239: 220–225.

Spratt BG, Staley JT, Fisher M (eds) (2005). Bacillus. In: Claridge MF, Dawah HW, Wilson MR (eds). Princeton Univ. Press: Princeton.

Staley JT (1997). Biodiversity: are microbial species threatened? Curr Opin Biotechnol 8: 340–345.

Staley JT (1997). Biodiversity: are microbial species threatened? Curr Opin Biotechnol 8: 340–345.

Staley JT (2006). The bacterial species dilemma and the genomic–phylogenetic species concept. Phil Trans R Soc London B 361: 1899–1909.

Staley JT, Castenholz RW, Colwell RR, Holt JK, Kane MD, Face NR et al. (1997). The Microbiol World: Foundation of the Biosphere. Am. Acad. Microbiol.: Washington DC.

Steunou A-S, Bhaya D, Bateson MM, Melendrez MC, Ward DM, Brecht E et al. (2006). In situ analysis of nitrogen fixation and
metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. *Proc Natl Acad Sci USA* **103**: 2398–2403.

Templeton AR (1989). The meaning of species and speciation: a genetic perspective. In: Otte D, Endler JA (eds). *Speciation and Its Consequences*. Sinauer: Sunderland, MA. pp 3–27.

Thompson JR, Pacocha S, Pharino C, Klepac-Ceraj V, Hunt DE, Benoit J et al. (2005). Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**: 1311–1313.

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW et al. (2005). Comparative metagenomics of microbial communities. *Science* **308**: 554–557.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM et al. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.

van der Meer MTJ, Schouten S, Bateson MM, Nübel U, Wieland A, Kühl M et al. (2005). Diel variations in carbon metabolisms of green nonsulfur-like bacteria in alkaline silicious hot spring microbial mats from Yellowstone National Park, USA. *Appl Environ Microbiol* **71**: 3978–3986.

van der Meer MTJ, Schouten S, Sinninghe Damsté JS, Ward DM (2007). Impact of carbon metabolisms on 13C signatures of cyanobacteria and green nonsulfur-like bacteria inhabiting a microbial mat from an alkaline silicious hot spring in Yellowstone National Park (USA) *Environ. Microbiology* **9**: 482–491.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Nature* **304**: 66–74.

Ward BB (2002). How many species of prokaryotes are there? *Proc Natl Acad Sci USA* **99**: 10234–10236.

Ward DM (1998). A natural species concept for prokaryotes. *Curr Opin Microbiol* **1**: 271–277.

Ward DM (2006). A macrobiological perspective on microbial species. *Microbe* **1**: 269–278.

Ward DM, Cohan FM (2005). Microbial diversity in hot spring cyanobacterial mats: pattern and prediction. In: Inskeep W, McDermott T (eds). *Geothermal Biology and Geochemistry in Yellowstone National Park*. Thermal Biology Institute, Montana State Univ: Bozeman. pp 185–201.

Ward DM, Bateson MM, Ferris MJ, Kühl M, Wieland A, Koeppe A et al. (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. *Phil Trans R Soc London B* **361**: 1997–2008.

Ward DM, Bateson MM, Ferris MJ, Nold SC (1998). A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiol Mol Biol Rev* **62**: 1353–1370.

Whitaker RJ, Banfield JF (2005). Population dynamics through the lens of extreme environments. *Rev Mineral Geochm* **59**: 259–277.

Whitaker RJ, Grogan DW, Taylor JW (2005). Recombination shapes the natural population structure of the hyperthermophilic archaeon *Sulfolobus islandicus*. *Mol Biol Evol* **22**: 2354–2361.

Zinser ER, Coe A, Johnson ZI, Martiny AC, Fuller NJ, Scanlan DJ et al. (2006). *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Appl Environ Microbiol* **72**: 723–732.