Review TRENDS in Microbiology Vol.15 No.9

Code and context: *Prochlorococcus* as a model for cross-scale biology

Maureen L. Coleman and Sallie W. Chisholm

Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Prochlorococcus **is a simple cyanobacterium that is abundant throughout large regions of the oceans, and has become a useful model for studying the nature and regulation of biological diversity across all scales of complexity. Recent work has revealed that environmental factors such as light, nutrients and predation influence diversity in different ways, changing our image of the structure and dynamics of the global *Prochlorococcus* population. Advances in metagenomics, transcription profiling and global ecosystem modeling promise to deliver an even greater understanding of this system and further demonstrate the power of cross-scale systems biology.**

The organism in context

Since the pioneering work on *Escherichia coli*, model organisms have been a cornerstone of molecular biology. Research on these model systems, however, rarely extends beyond the organism itself into the realm of the ecosystem in which it is embedded. Yet we know that the properties of all organisms are shaped by context, through adaptations to the heterogeneous, diverse and dynamic biotic and abiotic environments in which the organisms live. To understand living systems, we must understand their properties at all scales of organization, from the underlying pool of genetic information to the complex ecosystems that emerge from it.

New technologies are now enabling a cross-scale integrative biology, whereby a model organism can be investigated at scales from the molecular to the ecosystem, in the laboratory and *in situ*. This cross-scale biology seeks to elucidate reciprocal links between genes, metabolism, ecological interactions, environmental change and genome evolution. The requirements for such a system are demanding: it must be observable over time and space in nature and amenable to controlled growth under relevant conditions in the laboratory, and its natural habitat must be well characterized. The marine cyanobacterium *Prochlorococcus* meets these conditions (Box 1) and as such has advanced our understanding of microbial ecology and evolution.

Although the first evidence for the existence of *Prochlorococcus* dates back three decades [1,2], the isolation of *Prochlorococcus* into culture and the recognition of its significance as a globally abundant phototroph are more recent [3–6]. It is the numerically dominant photoautotroph at latitudes from 40°S to beyond 40°N in surface waters (upper 200 m) of the open ocean, making it one of the most abundant organisms on the planet [3] and an important object of study for oceanographers. The oligotrophic waters in which it thrives are characterized by steep gradients of light, temperature and nutrients, which vary not only with depth but also geographically and seasonally. The challenge is to understand how these environmental gradients dictate the distributions of ecological variants of *Prochlorococcus*, and how they drive genome evolution and diversification. Here we first describe how our understanding of light adaptation in *Prochlorococcus* has developed through studies at the cellular, genomic and population levels, both in cultured isolates and wild populations. We then apply a similar framework to understanding how nutrient availability and predators influence *Prochlorococcus* across scales.

The *Prochlorococcus* cell

Prochlorococcus is the smallest known oxygenic phototroph (0.5–0.7 μm diameter) and contains a unique photosynthetic apparatus. It is the only organism known to use divinyl chlorophyll *a* and *b* as the major light-harvesting pigments [7]. Furthermore, it harvests light with chlorophyll-binding antenna proteins (Pcb proteins) instead of the phycobilisomes used by most cyanobacteria, including its close relative, the marine *Synechococcus* [8].

Diversity viewed through the lens of light adaptation

Physiological differentiation

Prochlorococcus cells are found in abundance – typically 10^4–10^5 per ml – throughout the euphotic zone of the oceans, thriving at light intensities spanning four orders of magnitude. Isolates fall into two broad ecotypes (see Glossary) (Box 2) that are differentially adapted to high- and low-light conditions (HL and LL). HL cells can grow at

Glossary

Ecotype: a genetically and physiologically distinct population (see Box 2).
HLI: a phylogenetically distinct clade of high light-adapted *Prochlorococcus*, represented by the type strain MED4. This clade is also called 'Low B/A I' and 'eMED4'.
HLII: a phylogenetically distinct clade of high light-adapted *Prochlorococcus*, represented by the type strain MIT9312. This clade is also called 'Low B/A II' and 'eMIT9312'.
ITS: internal transcribed spacer sequence, located between the 16S and 23S rRNA genes.
LL: the low light-adapted *Prochlorococcus* clades. These clades are also referred to as 'High B/A' I–IV, and by their type strains: eNATL2A, eMIT9211, eSS120 and eMIT9313.

Corresponding author: Chisholm, S.W. (chisholm@mit.edu).
Available online 10 August 2007.

 0966-842X/$ – see front matter doi:10.1016/j.tim.2007.07.001

Box 1. *Prochlorococcus* as a model organism for cross-scale systems biology

Life is a system of systems, and thus can only be understood by studying it at all scales of biological organization. Each scale has properties that emerge from the interactions of its component parts, and those properties, in turn, influence the behavior of the component parts, and their evolution (Figure I).

Prochlorococcus has several features that make it a useful model system for elucidating the properties of life across scales, thereby connecting genomic information to global processes.

Genome features

- As an autotroph *Prochlorococcus* creates biomass from sunlight, CO_2 and inorganic nutrients, and thus has minimal requirements for growth.
- It has a small genome (roughly 2000 genes) and a simple regulatory system (as few as 28 predicted transcription regulators) [28,30].
- Many genome variants exist (12 cultured strains have been sequenced to date), which creates raw material for exploring evolutionary and functional diversity (genome sequences available in GenBank).

Cellular machinery and physiology

- *Prochlorococcus* is easily isolated into culture for studies of physiology, biochemistry and cellular systems biology [9,14], and its simple lifestyle – passively floating in a relatively well-mixed fluid medium – enables its natural environment to be recreated easily in the laboratory.
- Its limited requirements for growth greatly simplify physiological experiments and cellular modeling.

Population and community dynamics

- The natural habitat of *Prochlorococcus* is well studied, particularly at two long-term ocean time series sites near Hawaii and Bermuda (HOT and BATS) [68,69].
- Cells are abundant (up to 10^5 cells per ml), and can be easily identified and counted *in situ* [6].
- Wild cells can be easily sorted away from the rest of the microbial community using flow cytometry [14].
- Diverse strains of *Prochlorococcus*-infecting phage, which serve as a source of mortality and gene transfer, have been isolated [62].
- *Prochlorococcus* genes are abundant in metagenomic databases, enabling comparative genomics in the wild [42,63,64].

Global processes

- *Prochlorococcus* carries out a measurable fraction of global ocean production [3].
- Global models of its habitat are well developed, facilitating the simulation of its global dynamics and the relative fitness of different ecotypes in past, present and future oceans [65].

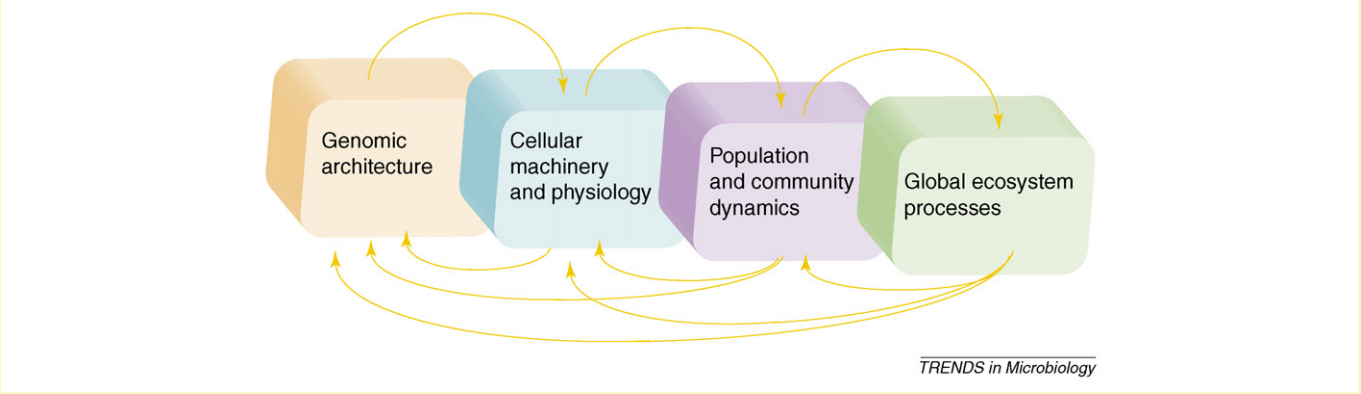

Figure I. Scales of biological organization. A comprehensive understanding of biological systems requires studying all scales of organization, and the reciprocal interactions across scales.

light intensities that are great enough to inhibit LL cells, whereas LL cells can grow at intensities that are too low to support the growth of HL strains [9]. Underlying these patterns are differences in light absorption efficiency resulting from different pigment ratios [8–10]. HL and LL cells also encode different forms of the phycobiliprotein phycoerythrin, although the role of this protein in light sensing or light harvesting is unknown [11,12]. This ecotype diversity helps to explain the broad habitat range of *Prochlorococcus*, and the ability of *Prochlorococcus* to dominate the base of the euphotic zone in the oceans. In fact, LL *Prochlorococcus* cells are arguably more efficient light harvesters than any other photosynthetic cell [13].

HL- and LL-adapted cells are distinguishable genetically at several loci [14–18], with HL cells forming a monophyletic clade divided into at least two subclades, HLI and HLII. By contrast, the LL isolates form at least four distinct lineages [16] (Figure 1a). These evolutionary reconstructions suggest that each clade might represent an ecologically distinct population, resulting from a selective sweep [16] (Box 2).

Distributions in the wild

Because these clades represent physiologically distinct cell types with different light optima for growth, we might expect their members to have different distributions along light (depth) gradients in the oceans. Indeed, initial field studies using the 16S rRNA and 16S–23S internal transcribed spacer (ITS) loci to differentiate between HL- and LL-adapted cells showed that although both coexist throughout the water column, HL cells outnumber LL cells in the surface by orders of magnitude. LL cells, although they usually have their abundance maximum in deeper waters, can be outnumbered by HL cells throughout the water column (Figure 1b) [19–24]. This asymmetry in distributions is probably the result of the asymmetry in physiological adaptation: HL cells have a greater fitness advantage, relative to LL cells, at high irradiances than do LL cells, relative to HL cells, at low irradiances [9].

As more data have accumulated we have learned that the situation is more complex. This classic pattern of HL

Box 2. What are ecotypes?

The term ecotype has been adopted in many different contexts to describe genetically and ecologically distinct units of diversity. We applied the term in describing physiological and genetic diversity in *Prochlorococcus* [14], and Cohan and others [70] have used it more formally in theoretical models of bacterial evolution. These two usages, however, are not (necessarily) equivalent.

In Cohan's stable ecotype model [70], a bacterial ecotype is a population of cells having the same ecological niche and whose divergence is constrained by periodic selection events. When one cell in a population acquires an adaptive mutation, the mutant and its descendants outcompete the rest of the population. Assuming that recombination is rare, the entire genome associated with the adaptive mutation will sweep through the population. Cohan proposes an approach for discovering ecotypes using sequence data, and under this stable ecotype model, sequence clusters will correspond one-to-one with ecotypes. One can then test whether these sequence-based ecotypes are indeed ecologically distinct.

Prochlorococcus ecotypes have been defined empirically using a combination of phenotypic and genetic data. High-light-adapted and low-light-adapted ecotypes were defined as coexisting populations with distinct photophysiology and phylogenetically distinct rRNA markers [14]. Thus, in contrast to the Cohan approach, the sequence clusters called ecotypes were chosen to reflect their known physiology. Within the HL ecotype, two ecologically distinct subclusters have been identified (see main text). Within the more divergent LL ecotype, several lineages are apparent. Although the term 'ecotype' has been applied to these LL lineages (i.e. eNATL2A, eSS120, eMIT9211 and eMIT9313 [19,20,26]), they are not well-resolved phylogenetically and their ecological differentiation is unclear.

Can we reconcile these two usages of the term 'ecotype'? If Cohan's stable ecotype model applies to *Prochlorococcus*, we should be able to delineate ecologically distinct populations as sequence clusters using a multilocus sequence analysis (MLSA) approach. In the absence of MLSA data, recent evidence hints that ecologically distinct lineages, distinguished by nutrient physiology, might exist within the named HL and LL *Prochlorococcus* clusters [34]. Several alternatives to the stable ecotype model predict multiple ecotypes within a sequence cluster, for instance when the recombination rate is high, when new traits are easily gained and lost by horizontal gene transfer, or when many ephemeral 'nano-niches' exist in the environment [70]. Testing these models in the *Prochlorococcus* system is an important step towards understanding the ecology and evolution of open-ocean bacterioplankton and developing a robust vocabulary to describe it.

cells dominating surface waters, and LL cells having an abundance maximum in deep waters, emerges when the water column has been stratified for some time. This physically isolates the surface and deep waters, enabling the HL and LL cells to photoacclimate and growth differences to manifest themselves in the population structure. When the water column is physically well mixed, however, LL cells can be as abundant as HL cells at the surface [25].

The two subclades of HL cells, HLI and HLII, display distinct distributions along oceanic gradients. HLII cells, represented by the type strain MIT9312, tend to be most abundant in warmer, highly stratified waters [19,20,23–27] and seem to be the most abundant cells on a global scale, whereas HLI cells (type strain MED4) dominate in cooler, weakly stratified waters [25,26] including the high-latitude cold-water limits of the *Prochlorococcus* range [26]. Consistent with these observations, laboratory studies have shown that HLI cells can grow at lower temperatures than HLII cells [26,27].

Among the LL-adapted clades, the forces shaping patterns of global distribution are less clear, partly because the diversity of LL lineages is not well represented in culture or in sequence databases [19,20]. When LL cells are found near the surface, they typically belong to one clade, represented by the cultured NATL2A strain, whereas another clade, represented by strain MIT9313, tends to localize deeper in the euphotic zone [19,20,23,26,27]. Although cultured isolates representing these two LL clades have similar light optima for growth, they might also possess different photoprotective mechanisms that would enable survival at elevated light intensity during mixing events [27]. Interestingly, this 'moderate' position of the NATL2A clade in the water column corresponds to its phylogenetic position as the most closely related LL clade to the HL clades [16].

Insights from whole genomes

The differentiation of HL and LL ecotypes is further revealed by whole-genome comparisons [28–30]. HL strain MED4, for example, has only one gene encoding the light-harvesting antenna protein Pcb, whereas LL strain MIT9313 has two, and LL strain SS120 – which grows at the lowest light intensity of these three [9] – has eight [31] (Figure 2). By contrast, MED4 has more genes encoding high-light-inducible proteins (HLIPs), thought to protect the cell from excess excitation energy [32], than do most LL strains (Figure 2). MED4 and other HL strains also encode photolyase, which repairs UV-induced DNA lesions, whereas the photolyase gene is absent from most LL strains [28–30]. Notably, however, the LL strains NATL1A and NATL2A possess a photolyase gene and 41 HLIPs (Figure 2), which might provide photoprotection and help explain the ability of this clade to thrive in high-light surface waters.

More surprising are the differences between the closely related HLI and HLII ecotypes. A genome comparison of two strains representing these two HL clades (MED4 and MIT9312) revealed that 10–15% of each genome is not found in the other strain, despite 99.2% 16S rRNA identity [33]. These differences are concentrated in a few major genomic islands, which contain signatures of horizontal gene transfer, including repeats, tRNA genes and genes that apparently came from phage genomes. Although the functional significance of island-encoded genes is unclear, many of them are expressed under certain conditions, suggesting that they are integrated into *Prochlorococcus* metabolic networks [33–35].

This analysis of light adaptation in *Prochlorococcus* serves as a template for beginning to understand the drivers of diversity in microorganisms. Through the simultaneous analysis of physiological properties of cultures, the distributions of their relatives in the wild, and their genome content, we can begin to tease apart the environmental factors that are most important in driving evolution and shaping contemporary populations.

Phosphorus economy and ecology

Cross-scale investigations are pointing to phosphorus (P) availability, which varies with both depth and geography, as a key ecological factor and a potential driver of genome

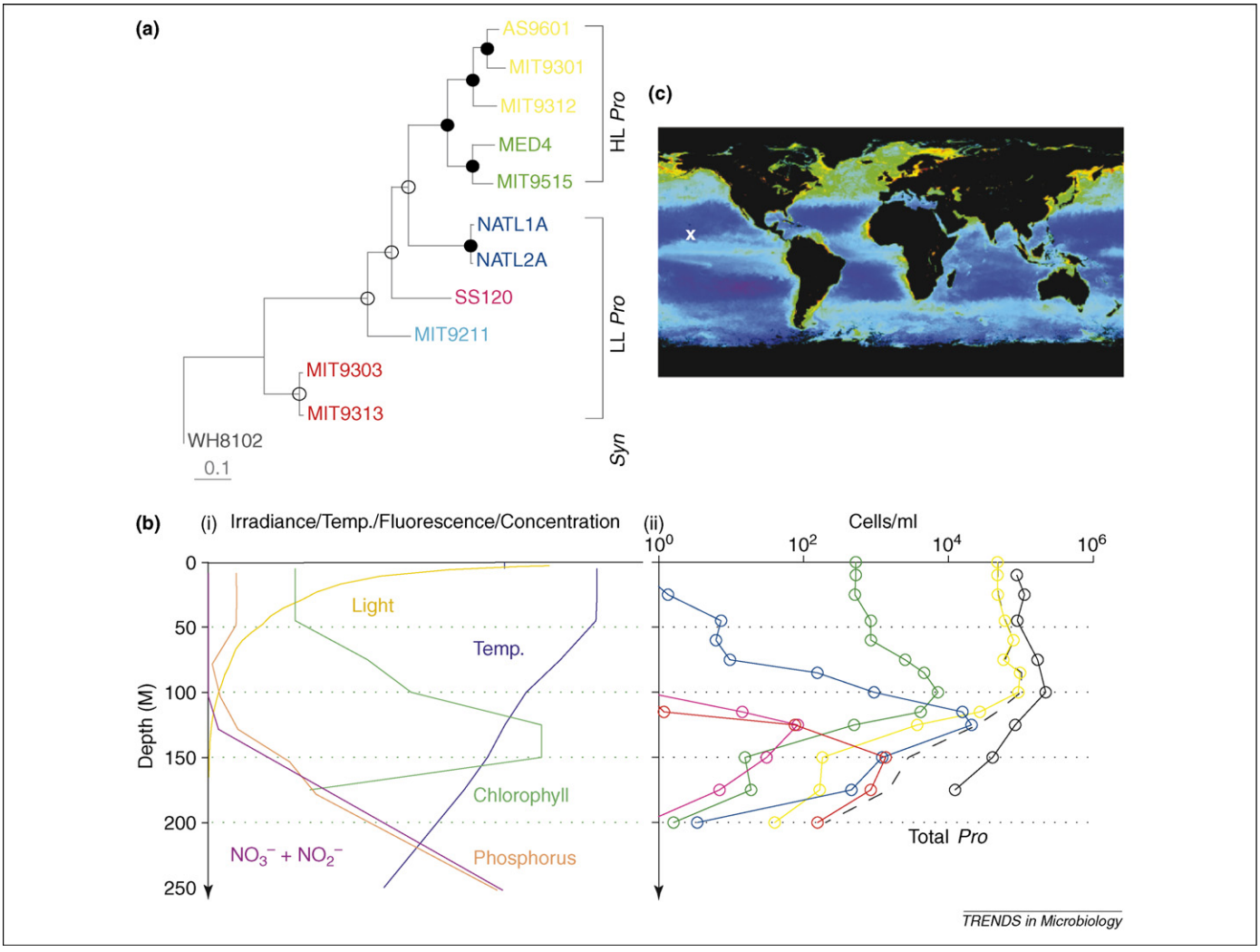

Figure 1. *Prochlorococcus* ecotypes: Evolutionary relationships and distributions along environmental gradients. **(a)** Phylogenetic tree showing the high-light-adapted and low-light-adapted clades (HL and LL), as defined by their physiological light-adaptation properties. The tree was generated from the *rpoB* gene by maximum likelihood (ML) with model parameters estimated from the data; nodes with >70% bootstrap support by two (unfilled circles) or three (filled circles) methods (distance, parsimony, ML) are shown. **(b)** (i) Depth profile (in meters) of water column characteristics [data from Hawaii Ocean Time-series (HOT): http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html] and (ii) abundance of *Prochlorococcus* clades (data from HOT; E. Zinser *et al.,* unpublished); both (i) and (ii) are from April 2003. Colors for each clade measured by quantitative PCR (qPCR). The black broken line is the sum of qPCR abundances, whereas the black unbroken line is total *Prochlorococcus* counted by flow cytometry. The discrepancy between the two totals indicates that current qPCR methods fail to detect some *Prochlorococcus* in deep waters. **(c)** Inset map shows surface water chlorophyll concentrations from NASA SeaWiFS (http://oceancolor.gsfc.nasa.gov/SeaWiFS/), seasonally averaged for spring 2003. HOT station shown by 'x'.

evolution in *Prochlorococcus*. This is not surprising, because P limits production in some regions of the oceans, with phosphate concentrations drawn down to the nanomolar range in some areas [36,37]. This extreme P depletion has selected for biochemical efficiency in *Prochlorococcus*, which has an unusually low cellular P:N ratio of 1:16–24 compared with other phytoplankton [38,39]. One biochemical adaptation enabling this P economy is the replacement of phospholipids with sulfolipids [40]. The small genome size (1.7 Mb for MED4) also reduces the P requirement relative to other essential elements, because the greatest demand for P in the cell is for nucleotide synthesis. Nevertheless, the MED4 chromosome contains more than half of the total cellular P [38].

Beyond this overall economization, different *Prochlorococcus* lineages possess further physiological adaptations for dealing with low P availability. Strain MED4 (HLI), for instance, can use a variety of organic P compounds for growth, whereas MIT9313 (LL) grows on a more limited range of P sources [41]. Furthermore, MED4 upregulates dozens of genes in response to P starvation, including known P-assimilation genes and novel genes of unknown function, which apparently enable it to endure prolonged P starvation [34]. By contrast, MIT9313 lacks many of these genes and is less capable of recovering from prolonged starvation [34]. These differences are in keeping with the preferred habitats of HL- and LL-adapted cells: surface waters often contain vanishingly small concentrations of inorganic phosphate, but regenerated organic P is available because of the intense biological activity in these waters. By contrast, LL *Prochlorococcus* cells are more abundant in deep waters, in closer proximity to the large reservoir of inorganic phosphate below the nutricline (Figure 1).

This simplistic HL–LL view of P acquisition quickly falls apart, however, as more strains are examined. MIT9312

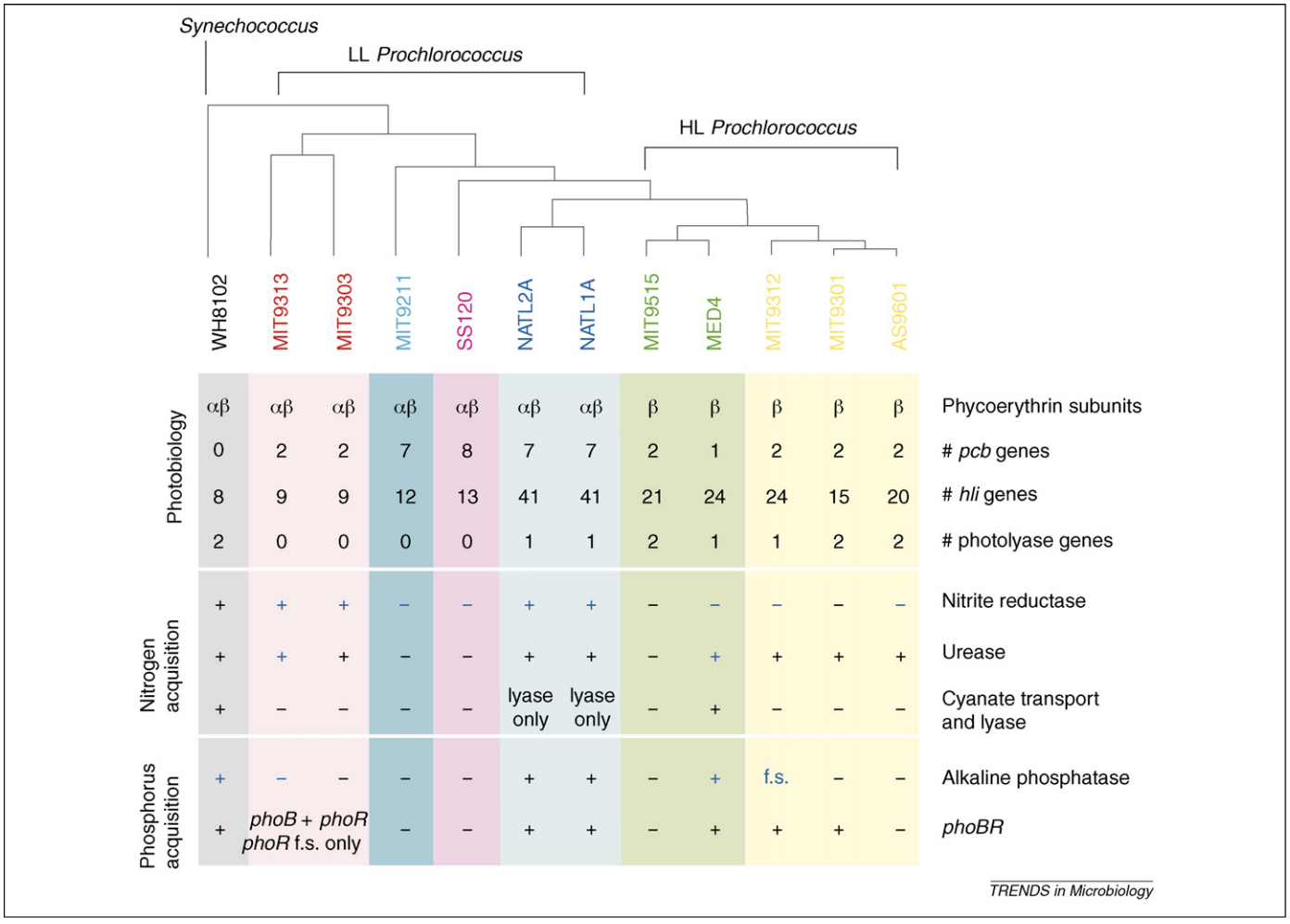

Figure 2. Presence or absence of genes involved in photosynthesis and nutrient acquisition in *Prochlorococcus* isolates. For +, – and f.s. entries in blue type, the activity of the gene, or lack thereof, has been confirmed by physiology experiments [41,45]. The presence of some genes, especially genes involved in photobiology, appears to map onto the phylogeny of the isolates, whereas the presence of nutrient acquisition genes varies even among the most closely related isolates. Cladogram shows the branching order of the isolates (from Figure 1). F.s. indicates frameshift mutation; + indicates presence of genes; – indicates absence of genes; and # indicates number of copies of a gene. Genomes from refs [28,30] and G. Kettler *et al.* (unpublished).

(HLII) lacks the broad P-utilization capabilities of MED4 [41]. In fact, among all HL isolates there is considerable variability in gene content related to P acquisition (Figure 2) [34]. In MED4, one large cluster of genes encodes regulation of P uptake (*phoBR*, *ptrA*), transport of phosphate (*phoE*, *pstSCAB*), and cleavage of organic P compounds (*phoA*), but nearly all of these genes are missing from HLI strain MIT9515, which shares 99.9% 16S rRNA identity with MED4. Similarly, HLII strains AS9601, MIT9301 and MIT9312 share 99.9% 16S rRNA identity, but they possess different gene complements for P acquisition. This variability in gene content is thought to reflect variation in selection imposed by P availability in the oceans: isolates from P-limited regions such as the Mediterranean and Sargasso Sea (e.g. MED4, MIT9301) have more genes for P acquisition, whereas isolates from iron-limited regions like the equatorial Pacific (e.g. MIT9515), or isolates from deep in the water column (e.g. MIT9312, MIT9313), have fewer genes for P acquisition [34]. Thus in addition to the vertical (depth) component there seems to be a horizontal or geographic component to P adaptation.

This hypothesis is supported by recent metagenomic data showing that the abundance of phosphate transport genes is correlated with P availability [42]. In the low-P waters of the Caribbean, for instance, the gene encoding the phosphate-binding protein PstS is seven times more abundant compared with the higher-P waters of the Pacific. Several P-related genes, including *pstS* and *phoH*, are also found in cyanophage genomes, suggesting that P might limit phage replication and offering a potential mechanism for lateral gene transfer of P genes among hosts [43]. An important next step is to combine these genome surveys with *in situ* functional measurements of gene expression, protein abundance and activity [24], and phosphate uptake rates, to understand how these gene content changes translate to physiology.

Obtaining nitrogen: more than one solution

Prochlorococcus does not fix dinitrogen gas (N_2), and surprisingly none of the cultured isolates can use nitrate as a nitrogen (N) source. Nevertheless, the upper waters of the open ocean contain several other N sources, and *Prochlorococcus* seems to use several of them. As expected, all strains can use ammonium, the form incorporated biosynthetically by glutamine synthetase and generated through recycling. Some, but not all, LL strains possess

nitrite reductase (Figure 2) and can use nitrite as their sole N source, consistent with the fact that their optimal light intensity for growth often occurs near the depth of the nitrite maximum in the ocean water column. Interestingly, the nitrite permease gene in MIT9313 seems to have been horizontally transferred [28]. Likewise horizontal transfer might have introduced a cyanate transporter and lyase in MED4 and an amino acid transporter in MIT9312, both in genomic islands [33]. Field studies suggest that *Prochlorococcus* can take up amino acids at an elevated rate [44], suggesting that these genes might be widespread in the wild. Finally, several HL and LL strains can grow on urea [45] (Figure 2).

The inability of cultured *Prochlorococcus* to grow on nitrate is surprising, given that virtually all cultured *Synechococcus*, a close relative, seem to have this capability and that nitrate can be abundant in the deep euphotic zone. Nitrate must be reduced to ammonium for biosynthesis, however, and it is possible that energy limitation precludes this pathway at extreme depths. We suspect that directed isolation approaches will yield nitrate-using *Prochlorococcus* from environments with sufficient nitrate and light.

In addition to having altered gene content, *Prochlorococcus* lineages seem to have adapted to different N regimes by altering the regulation of core genes. For example, the P_{II} protein (encoded by *glnB)*, which coordinates carbon and nitrogen metabolism, responds differently to N starvation in HL strain MED4 and LL strain MIT9313 [35]. The regulation of N metabolism in *Prochlorococcus* seems to be simpler than in other cyanobacteria, probably as a result of the unique selective pressures in oligotrophic environments [46].

Behind the scenes: essential micronutrients

In the open oceans, vanishingly low trace metal concentrations are maintained by, and select for, efficient uptake mechanisms in microorganisms [47]. These trace metals are essential cofactors for the metalloenzymes underlying major metabolic processes such as photosynthesis, carbon fixation and nutrient assimilation. Cobalt [48] and nickel (for superoxide dismutase [49]) are required for *Prochlorococcus* growth, and copper is probably required for plastocyanin. But copper can also be toxic to *Prochlorococcus*, even at the low concentrations found in the surface waters of oligotrophic oceans. HL cells are more resistant to copper toxicity than LL cells, consistent with the observation that free copper concentrations are greater near the surface [50]. Given this concentration gradient with depth, and the role of metals in both causing and curing (e.g. through superoxide dismutase) oxidative stress, we suspect that requirements for and sensitivities to trace metals will frequently be related to the light adaptation of the strain.

Iron has a key role in photophysiology and thus we also expect to find different iron physiologies in HL and LL ecotypes of *Prochlorococcus*. Iron is required in large amounts for photosystems but occurs at low bioavailable concentrations throughout much of the open oceans, and as a result limits primary production in these regions [51]. Indeed, the cell division rate, cell size and cellular chlorophyll content of *Prochlorococcus* have been shown to increase in response to iron addition in the equatorial Pacific, indicating cellular iron limitation [52,53]. The physiological iron–light interaction is exemplified by the light-harvesting antenna proteins (Pcb proteins). When iron is limiting, the LL strains SS120 and MIT9313, but not the HL strain MED4, upregulate specific Pcb antenna proteins that might be necessary for light harvesting or photoprotection (or both) [54]. The demand for iron, particularly by cells at low light intensity that need more photosystems, has undoubtedly influenced the genomes and physiology of *Prochlorococcus* in other ways that remain to be elucidated.

Pressure from the top

The net growth rate and population size of *Prochlorococcus* in the wild depends not only on resource availability, which limits the cell division rate, but also on mortality. Predation seems to balance *Prochlorococcus* growth rate on average [55,56]. Several studies suggest that although nutrient limitation dictates cell division rate, cell abundance is not a simple function of nutrient concentrations, because grazing rate also varies with nutrients [52,55]. Thus predators might regulate *Prochlorococcus* biomass directly whereas nutrients might select for subsets of the population with the most efficient acquisition and most robust stress responses. Predators might also influence population structure by selecting for certain cell types based on, for instance, cell surface properties [57].

Although viruses are now recognized as being abundant in the oceans, their contribution to bacterial mortality is unclear; estimates of virus-induced mortality range from 1% to 50% [58]. It is clear, however, that phages are important for *Prochlorococcus* evolution and diversity. Photosynthesis genes are found in phage genomes and probably recombine with host versions, thereby increasing diversity [59–61]. Phage genomes also contain genes involved in phosphate and carbon metabolism, suggesting further coevolution of host and phage [43]. Furthermore, *Prochlorococcus* genomes harbor strain-specific genomic islands encoding biosynthesis of lipopolysaccharides and other cell surface features [28,33]; these islands might help explain the different phage susceptibility [62] observed for cultured isolates, and might contribute to differential mortality and relative fitness of cells in the wild.

Revealing organization and complexity through metagenomics

Recently there has been a flood of metagenomics data from *Prochlorococcus*-rich regions of the oceans [42,63,64], and these data offer an unprecedented look at evolutionary processes and patterns of diversity. For example, large-insert fosmid sequences can reveal insertions and deletions of genes in wild cells compared with cultured isolates, whereas small-insert shotgun libraries can reveal large-scale patterns of gene presence and absence in the whole community (Figure 3). These datasets also reveal that the extensive sequence diversity is clearly structured, and that regions of the genome are hypervariable even in a single population [42]. Moreover, abundant cyanophage sequences in the cellular size fraction, probably derived from replicating intracellular phages, reinforce the importance

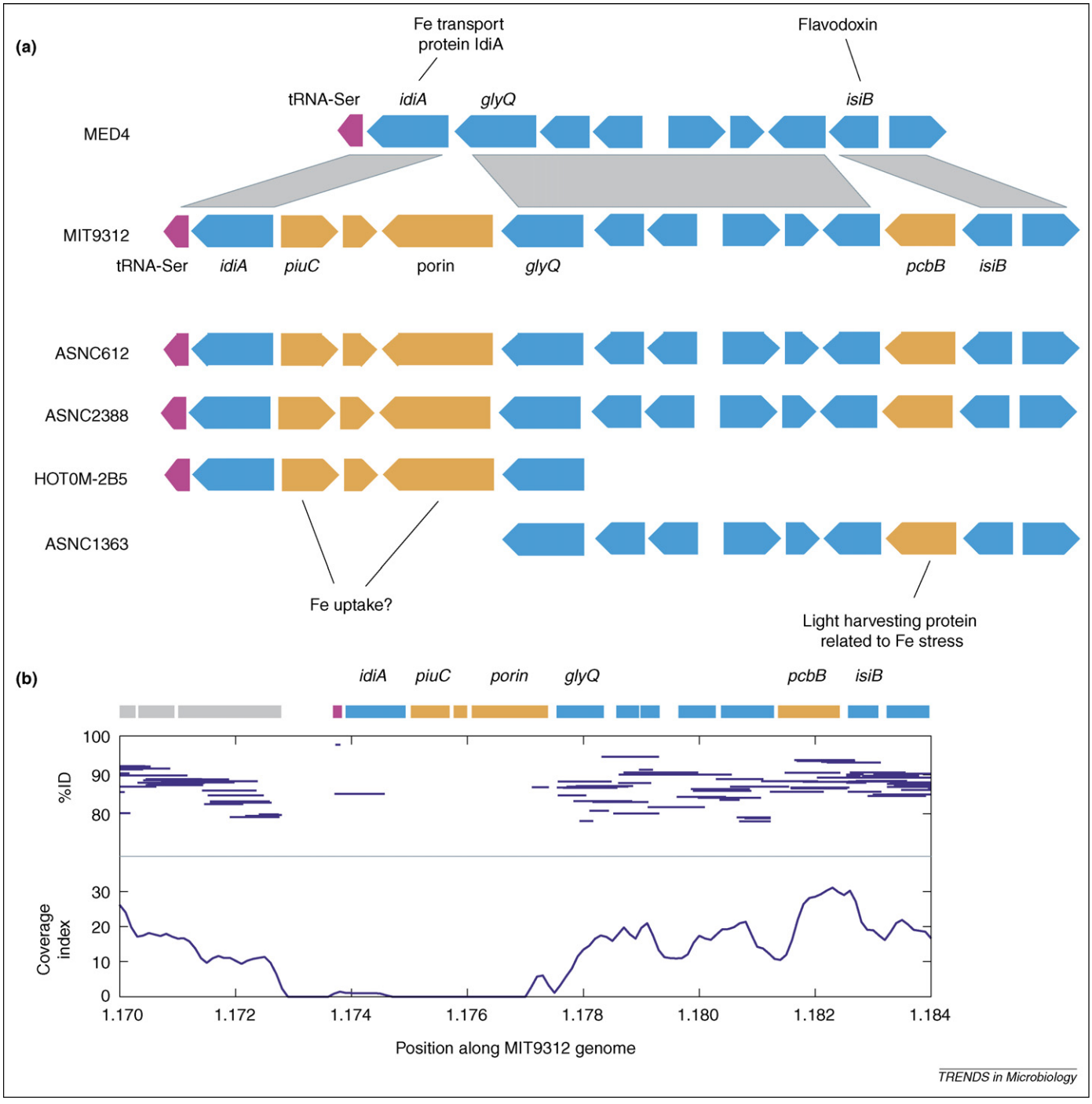

Figure 3. Patterns of diversity in wild *Prochlorococcus* revealed through metagenomics. Here we use metagenomics data to test whether genome variation observed in cultured strains of *Prochlorococcus* is observed in wild populations from two different (Atlantic and Pacific) oceanic regimes. **(a)** A genome fragment from two cultured strains, MED4 and MIT9312, compared with the homologous section of four large pieces of *Prochlorococcus* DNA collected from the Hawaii Ocean Time-series station (HOT) [33,64]. This region of the genome contains several genes thought to be involved in iron stress [28]. MED4 lacks the genes identified with orange color, whereas these particular wild cells from HOT all have them. **(b)** Smaller genome fragments from cells from the Sargasso Sea [63] aligned to the MIT9312 genome in this same region. The top panel shows the percent nucleotide identity (%ID) between the fragments and homologous regions in the MIT9312 genome, revealing ~80–90% identity in shared genes. The bottom panel shows an estimate of 'coverage', or how many sequences in this sample were homologous to a particular section of the genome. The Sargasso population seems to lack several putative iron stress genes found in the Pacific.

of these viruses for mortality and host evolution [64]. Metagenomic data have been so informative for *Prochlorococcus* and its viruses in part because of the availability of whole-genome sequences of cultured isolates, which serve as scaffolds for interpreting the environmental data. Thus it is crucial that future sequencing efforts include both wild populations and whole genomes of cultured isolates.

Conclusions and future perspectives

A picture of *Prochlorococcus* ecology and evolution is emerging thanks to the combination of approaches and scales of interrogation employed (Box 1; Figure 4). Clearly this picture is far more complex than can be captured by the HL–LL ecotype paradigm that first emerged from studies of light physiology and molecular phylogenies. For nutrients like N and P, gene loss and gene gain have occurred

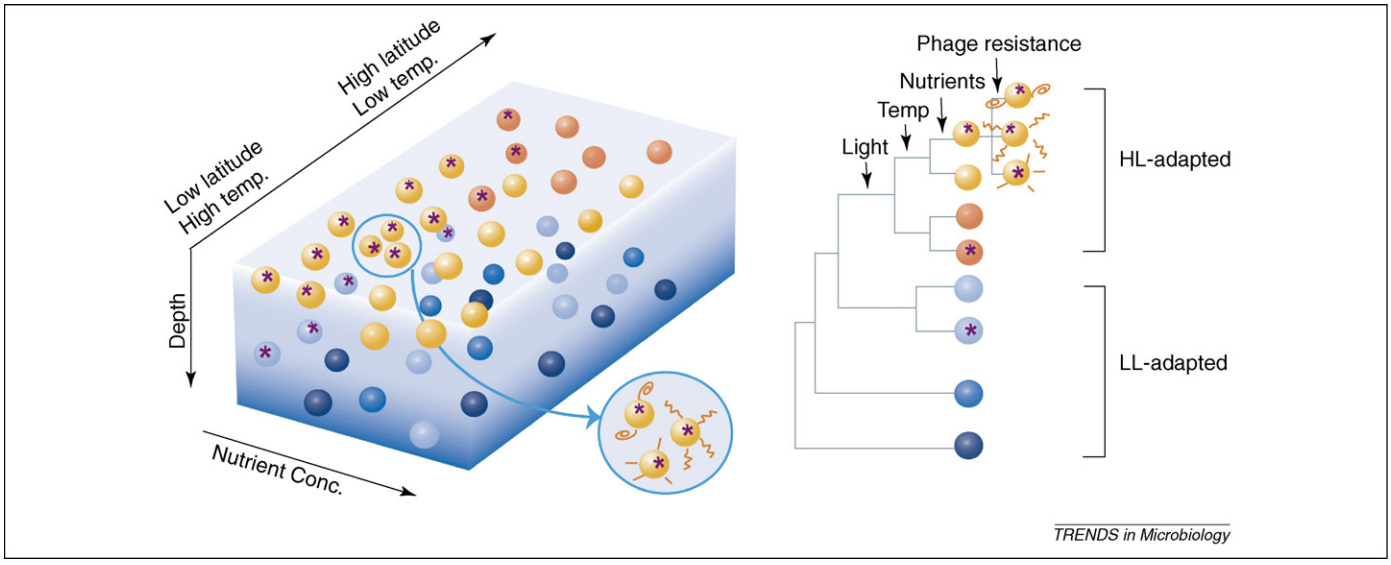

Figure 4. Hypothetical patterns of diversity along different *Prochlorococcus* niche dimensions and spatial scales. Within the globally distributed HL-adapted group, we might see differentiation based on temperature, and within each temperature-adapted clade we might see multiple 'nutrient ecotypes' at more local scales, and multiple phage-resistance types within a single nutrient ecotype. A robust understanding of these patterns will require all of the tools at our disposal: metagenomics, single cell genomics and models of cellular metabolism. Key: spheres represent *Prochlorococcus* cells. Shades of orange represent HL-adapted clades; shades of blue, LL-adapted clades. Purple stars represent additional adaptations to a specific nutrient-depleted environment (e.g. low P, low Fe). Exterior decoration represents cell surface diversity, such as different lipopolysaccharide structures or surface proteins, which might confer resistance to certain phages or grazers.

within the 'ecotypes' delineated by rRNA ITS sequence clusters, and as a result nutrient physiology does not map neatly onto the ITS phylogeny. At an even finer resolution, extremely closely related cells have different phage susceptibilities (Figure 4). Thus, the recognition of clades and clusters, and their interpretation in light of ecological factors, depends on the scale of observation.

Prochlorococcus is but one of many model organisms (e.g. *Pelagibacter*, *Roseobacter*, *Vibrio* and *Synechococcus*) that are proving useful for understanding marine microbial processes, and biological systems in general. As a microbially dominated ecosystem, where the physics and chemistry are fairly well understood and modeled, the oceans provide us with an opportunity to connect, for the first time, the information in genomes to the dynamics of whole ecosystems [65]. One of the new frontiers for marine microbiology is moving from the 'parts list', that is, the genetic information carried by marine microbes, to understanding the function of the metabolic machinery of the oceans. This machinery is an integral component of global biogeochemical cycles, and we must strive to understand how it responds to perturbations at local and global scales. Techniques for measuring gene expression *in situ* at both the RNA and protein levels are on the horizon, and these can be validated using model systems like *Prochlorococcus*. Furthermore, single cell analyses [66,67] combined with revolutionary advances in DNA sequencing technologies will change the way we study marine microbial systems.

Even in these relatively simple living systems, we are humbled and challenged by the complexity and robustness of nature. Our hope is that *Prochlorococcus* and other model microbial systems from the oceans will not only advance our understanding of microbial processes, but also yield fundamental insights into the structure, function and evolution of life in general.

Acknowledgements

We thank members of the Chisholm Laboratory, especially J. Waldbauer and M. Sullivan, for comments on the manuscript: Ed DeLong for providing fosmid clones; and *Prochlorococcus* enthusiasts, past and present, for their contributions to this story. We thank Jim Tiedje, Dan Drell, Frank Larimer, Elbert Branscomb and the DOE Joint Genome Institute for guiding the entry of *Prochlorococcus* into the genomic era. This research on *Prochlorococcus* has been supported over the years by grants from the DOE, NSF, ONR and the Seaver Foundation (to S.W.C.). M.L.C. is supported by an NSF Graduate Fellowship. Particular thanks are due to the Gordon and Betty Moore Foundation for their support of some of this research, and their recent transformative role in the field of marine microbiology.

References

1 Johnson, P.W. and Sieburth, J.M. (1979) Chroococcoid cyanobacteria in the sea – ubiquitous and diverse phototropic biomass. *Limnol. Oceanogr.* 24, 928–935
2 Gieskes, W.W. and Kraay, G.W. (1983) Unknown chlorophyll-a derivatives in the North-Sea and the tropical Atlantic-Ocean revealed by HPLC analysis. *Limnol. Oceanogr.* 28, 757–766
3 Partensky, F. *et al.* (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* 63, 106–127
4 Chisholm, S.W. *et al.* (1992) *Prochlorococcus-marinus* nov gen-nov sp – an oxyphototrophic marine prokaryote containing divinyl chlorophyll-a and chlorophyll-B. *Arch. Microbiol.* 157, 297–300
5 Chisholm, S.W. *et al.* (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* 334, 340–343
6 Olson, R.J. *et al.* (1990) Spatial and temporal distributions of prochlorophyte picoplankton in the North-Atlantic Ocean. *Deep-Sea Res.* 37, 1033–1051
7 Goericke, R. and Repeta, D.J. (1992) The pigments of *Prochlorococcus marinus* – the presence of divinyl chlorophyll-a and chlorophyll-B in a marine prokaryote. *Limnol. Oceanogr.* 37, 425–433
8 Partensky, F. *et al.* (1997) The divinyl-chlorophyll a/b-protein complexes of two strains of the oxyphototrophic marine prokaryote *Prochlorococcus* – Characterization and response to changes in growth irradiance. *Photosynth. Res.* 51, 209–222
9 Moore, L.R. and Chisholm, S.W. (1999) Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol. Oceanogr.* 44, 628–638

10 Moore, L.R. *et al.* (1995) Comparative physiology of *Synechococcus* and *Prochlorococcus* – Influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Mar. Ecol. Prog. Ser.* 116, 259–275
11 Steglich, C. *et al.* (2005) A green light-absorbing phycoerythrin is present in the high-light-adapted marine cyanobacterium *Prochlorococcus* sp MED4. *Environ. Microbiol.* 7, 1611–1618
12 Steglich, C. *et al.* (2003) Photophysical properties of *Prochlorococcus marinus* SS120 divinyl chlorophylls and phycoerythrin *in vitro* and *in vivo*. *FEBS Lett.* 553, 79–84
13 Morel, A. *et al.* (1993) *Prochlorococcus* and *Synechococcus* – a comparative study of their optical properties in relation to their size and pigmentation. *J. Mar. Res.* 51, 617–649
14 Moore, L.R. *et al.* (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* 393, 464–467
15 Urbach, E. *et al.* (1998) Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria). *J. Mol. Evol.* 46, 188–201
16 Rocap, G. *et al.* (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl. Environ. Microbiol.* 68, 1180–1191
17 Ferris, M.J. and Palenik, B. (1998) Niche adaptation in ocean cyanobacteria. *Nature* 396, 226–228
18 Steglich, C. *et al.* (2003) Analysis of natural populations of *Prochlorococcus* spp. in the northern Red Sea using phycoerythrin gene sequences. *Environ. Microbiol.* 5, 681–690
19 Ahlgren, N.A. *et al.* (2006) Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environ. Microbiol.* 8, 441–454
20 Zinser, E.R. *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
21 West, N.J. and Scanlan, D.J. (1999) Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Appl. Environ. Microbiol.* 65, 2585–2591
22 West, N.J. *et al.* (2001) Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by *in situ* hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology* 147, 1731–1744
23 Fuller, N.J. *et al.* (2006) Molecular analysis of picocyanobacterial community structure along an Arabian Sea transect reveals distinct spatial separation of lineages. *Limnol. Oceanogr.* 51, 2515–2526
24 Fuller, N.J. *et al.* (2005) Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. *Limnol. Oceanogr.* 50, 363–375
25 Bouman, H.A. *et al.* (2006) Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* 312, 918–921
26 Johnson, Z.I. *et al.* (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* 311, 1737–1740
27 Zinser, E.R. *et al.* Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol. Oceanogr.* (in press)
28 Rocap, G. *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424, 1042–1047
29 Hess, W.R. *et al.* (2001) The photosynthetic apparatus of *Prochlorococcus*: insights through comparative genomics. *Photosynth. Res.* 70, 53–71
30 Dufresne, A. *et al.* (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10020–10025
31 Garczarek, L. *et al.* (2000) Multiplication of antenna genes as a major adaptation to low light in a marine prokaryote. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4098–4101
32 He, Q. *et al.* (2001) The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. *J. Biol. Chem.* 276, 306–314
33 Coleman, M.L. *et al.* (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* 311, 1768–1770
34 Martiny, A.C. *et al.* (2006) Phosphate acquisition genes in *Prochlorococcus* ecotpes: evidence for genome-wide adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12552–12557
35 Tolonen, A.C. *et al.* (2006) Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol. Syst. Biol.* 2, 53
36 Ammerman, J.W. *et al.* (2003) Phosphorus deficiency in the Atlantic: an emerging paradigm in oceanography. *Eos. Trans. A.G.U.* 84, 165
37 Thingstad, T.F. *et al.* (2005) Nature of phosphorus limitation in the ultraoligotrophic eastern Mediterranean. *Science* 309, 1068–1071
38 Bertilsson, S. *et al.* (2003) Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea. *Limnol. Oceanogr.* 48, 1721–1731
39 Heldal, M. *et al.* (2003) Elemental composition of single cells of various strains of marine *Prochlorococcus* and *Synechococcus* using X-ray microanalysis. *Limnol. Oceanogr.* 48, 1732–1743
40 Van Mooy, B.A. *et al.* (2006) Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8607–8612
41 Moore, L.R. *et al.* (2005) Ecotypic variation in phosphorus-acquisition mechanisms within marine picocyanobacteria. *Aquat. Microb. Ecol.* 39, 257–269
42 Rusch, D.B. *et al.* (2007) The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* 5, e77
43 Sullivan, M.B. *et al.* (2005) Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* 3, e144
44 Zubkov, M.V. *et al.* (2003) High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl. Environ. Microbiol.* 69, 1299–1304
45 Moore, L.R. *et al.* (2002) Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnol. Oceanogr.* 47, 989–996
46 Garcia-Fernandez, J.M. *et al.* (2004) Streamlined regulation and gene loss as adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic environments. *Microbiol. Mol. Biol. Rev.* 68, 630
47 Morel, F.M.M. and Price, N.M. (2003) The biogeochemical cycles of trace metals in the oceans. *Science* 300, 944–947
48 Saito, M.A. *et al.* (2002) Cobalt limitation and uptake in *Prochlorococcus*. *Limnol. Oceanogr.* 47, 1629–1636
49 Eitinger, T. (2004) *In vivo* production of active nickel superoxide dismutase from *Prochlorococcus marinus* MIT9313 is dependent on its cognate peptidase. *J. Bacteriol.* 186, 7821–7825
50 Mann, E.L. *et al.* (2002) Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnol. Oceanogr.* 47, 976–988
51 Boyd, P.W. *et al.* (2007) Mesoscale iron enrichment experiments 1993-2005: Synthesis and future directions. *Science* 315, 612–617
52 Mann, E.L. and Chisholm, S.W. (2000) Iron limits the cell division rate of *Prochlorococcus* in the eastern equatorial Pacific. *Limnol. Oceanogr.* 45, 1067–1076
53 Cavender-Bares, K.K. *et al.* (1999) Differential response of equatorial Pacific phytoplankton to iron fertilization. *Limnol. Oceanogr.* 44, 237–246
54 Bibby, T.S. *et al.* (2003) Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem. *Nature* 424, 1051–1054
55 Worden, A.Z. and Binder, B.J. (2003) Application of dilution experiments for measuring growth and mortality rates among *Prochlorococcus* and *Synechococcus* populations in oligotrophic environments. *Aquat. Microb. Ecol.* 30, 159–174
56 Landry, M.R. *et al.* (2003) Phytoplankton growth and microzooplankton grazing in high-nutrient, low-chlorophyll waters of the equatorial Pacific: Community and taxon-specific rate assessments from pigment and flow cytometric analyses. *J. Geophys. Res.* 108, 8142
57 Monger, B.C. *et al.* (1999) Feeding selection of heterotrophic marine nanoflagellates based on the surface hydrophobicity of their picoplankton prey. *Limnol. Oceanogr.* 44, 1917–1927
58 Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548

59 Lindell, D. *et al.* (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11013–11018
60 Sullivan, M.B. *et al.* (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* 4, 1344–1357
61 Zeidner, G. *et al.* (2005) Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environ. Microbiol.* 7, 1505–1513
62 Sullivan, M.B. *et al.* (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* 424, 1047–1051
63 Venter, J.C. *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74
64 DeLong, E.F. *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311, 496–503
65 Follows, M. *et al.* (2007) Emergent biogeography of microbial communities in a model ocean. *Science* 315, 1843–1846
66 Zhang, K. *et al.* (2006) Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.* 24, 680–686
67 Stepanauskas, R. and Sieracki, M.E. (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc. Natl. Acad. Sci. U. S. A.* 104, 9052–9057
68 DuRand, M.D. *et al.* (2001) Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep-Sea Res. Pt II* 48, 1983–2003
69 Campbell, L. *et al.* (1994) The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. *Limnol. Oceanogr.* 39, 954–961
70 Cohan, F.M. (2006) Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philos. Trans. R. Soc.Lond. B Biol. Sci.* 361, 1985–1996