Prochlorococcus: the structure and function of collective diversity

Prochlorococcus: The structure and function of collective diversity

Steven J. Biller[1], Paul M. Berube[1], Debbie Lindell[2], Sallie W. Chisholm[1,3]

[1]Department of Civil and Environmental Engineering & [3]Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139 USA.
[2]Faculty of Biology, Technion – Israel Institute of Technology, Haifa 32000, Israel.

Correspondence to SWC. E-mail: chisholm@mit.edu

The marine cyanobacterium *Prochlorococcus* is the smallest and most abundant photosynthetic organism on Earth. In this Review, we synthesize our understanding of the diversity of this remarkable phototroph and describe its role in ocean ecosystems. We discuss the importance of interactions with the physical environment, phages and heterotrophs in shaping the ecology and evolution of this group. In light of recent studies, we have come to view *Prochlorococcus* as a federation of microdiverse cells that sustains its broad distribution, stability and abundance in the oceans through extensive genomic and phenotypic diversity. Thus, it is proving to be a useful model system for elucidating the forces that shape microbial populations and ecosystems.

Since the discovery of *Prochlorococcus* in 1985[1], considerable progress has been made in understanding the characteristics that make this cell unique in the microbial world. It is the smallest (0.5 – 0.7 μm diameter)[2] and most abundant photosynthetic cell on the planet, with an estimated global population of ~10^{27} cells[3-5]. *Prochlorococcus* has the smallest genome of any free-living phototroph[6]; some isolates have genomes as small as 1.65 Mbp, with only ~1700 genes[7]. It is the only marine phytoplankter that uses the divinyl form of chlorophylls *a* and *b* to harvest light energy[8], which causes a slight red-shift in their absorption spectra[2,9]. This unique pigmentation makes it possible to determine that *Prochlorococcus* accounts for 50% of the total chlorophyll in vast stretches of the surface oceans[10,11]. Collectively, this cyanobacterium produces an estimated 4 gigatons of fixed carbon each year[3] – approximately the same net primary productivity as global croplands[12].

Prochlorococcus thrives throughout the euphotic zone of the tropical and subtropical oligotrophic ocean[1,13]. The daily light-dark cycle synchronizes cell division in *Prochlorococcus*[14] and is an important driver of highly choreographed gene expression patterns throughout the day[15-19]. The euphotic zone is shaped by continuous macro-scale gradients of light, temperature, and nutrients (Fig 1A). Both light intensity and temperature are highest at the surface and decrease with depth, whereas nutrient levels are typically low at the surface and increase gradually with depth. Although there is fine-scale variation within the water column, the physical and chemical environment of the ocean as a whole tends to be constrained and less variable than many other microbial habitats such as soils, and exhibits gradual changes on monthly to annual timescales.

From the perspective of its microbial inhabitants, the oligotrophic ocean is an extremely dilute environment both in terms of its chemistry and biology (Fig 1B). For example, the average *Prochlorococcus* cell may be hundreds of cell diameters away from another cell of any type, and even a few cell diameters away from essential nutrients, which are found at pico- to nanomolar concentrations. By contrast, well-studied model microorganisms, such as *Escherichia coli*, tend to reside in relatively nutrient-rich and densely populated environments, such as the gut. To overcome the challenges associated with their dilute environment, studies have shown that some marine microorganisms attach to particles[20] or form close associations with other bacteria[21]. Although microscale patchiness of some form may contribute to the survival of *Prochlorococcus*, direct evidence to support or reject this hypothesis is lacking. Nevertheless, the dilute nature of oligotrophic ecosystems clearly imposes a unique set of selective pressures on microbial life.

Prochlorococcus has several traits that make it well suited to this dilute habitat. Compared to other phytoplankton, *Prochlorococcus* cells have a low P requirement (they have high C:P and N:P ratios)[22-24], partly due to their relatively small genomes[22] and the substitution

of sulpholipids for phospholipids in the cell membrane[25]. Their small size results in a high surface to volume ratio that facilitates efficient nutrient acquisition and enhances light absorption which, when combined with their unique pigmentation[9], make them the most efficient light absorbers of any photosynthetic cell; they are the only phytoplankton known to absorb more light than they scatter[2]. Thus, *Prochlorococcus* can thrive at lower light intensities than those required by most other phytoplankton[9] and their populations extend deeper in the water column than almost any other phototroph[26] – essentially defining the lower boundary of photosynthetic life in the oceans.

The ability of *Prochlorococcus* to occupy the entire euphotic zone can be explained largely by its microdiversity, with different subgroups adapted to different light optima for growth[9,27]. Strains isolated from deep waters grow optimally at substantially lower light intensities (termed low light (LL)-adapted ecotypes) than those isolated from the surface (termed high light (HL)-adapted ecotypes) (Fig. 1C). These adaptations result in niche-partitioning in the water column[28-30]: HL-adapted cells are orders of magnitude more abundant in surface waters[31-33], but are outnumbered by LL-adapted cells at the base of the euphotic zone (Fig 1C). Despite the complexity of ocean dynamics, these distinct groups of *Prochlorococcus* shift in relative abundance in reproducible annual cycles[34,35], demonstrating the remarkable robustness of *Prochlorococcus* populations. The distinction between HL and LL-adapted cells forms the basis of our understanding of *Prochlorococcus* diversity, but, as discussed in this Review, light is just one of many factors that has driven the diversification of this bacterial group[36]. Here, we discuss the genomic diversity of *Prochlorococcus*, the factors that contribute to this diversity and its consequences for the ecology of this marine cyanobacterium**.**

Deeply rooted evolutionary diversity

The 16S rRNA sequences of all *Prochlorococcus* isolates do not differ by more than ~3% – the traditional boundary for defining a microbial species. Thus, this group has maintained a coherent identity, even though it exhibits extraordinary diversity in other traits (see below). Given the conservation of 16S rRNA, the internal transcribed spacer (ITS) sequence between the 16S and 23S sequence is typically used to provide increased phylogenetic resolution[37]. Many other marker genes (including *rpoC1*[38,39], *petB/D*[40], *ntcA*[41], and *gyrB*[42]), have been used as well, and provide similar insight into its evolutionary history. *Prochlorococcus* is a monophyletic group that is closely related to marine *Synechococcus*[28,37,40,43], although key physiological and ecological features distinguish the two genera (Box 1). ITS-based trees of *Prochlorococcus* are surprisingly consistent with those derived from whole genome protein coding sequences, making this a useful marker for exploring evolutionary relationships[44,45].

Although the initial partitioning of *Prochlorococcus* into the broad categories of HL and LL-adapted strains was based solely on phenotype[9,27,46], molecular phylogenic analyses have revealed that this division corresponds with the earliest phylogenetic split within the *Prochlorococcus* lineage[31,37,38,40]. HL-adapted *Prochlorococcus* strains form a coherent group that resolves into at least 6 clades (HLI-HLVI) (Fig 2A)[9,27,31,46], whereas the polyphyletic LL-adapted strains partition into at least seven clades (LLI-LLVII)[36,47,48]. Although alternative nomenclature has been proposed, the HL and LL notation has emerged as the most consistent way to refer to the ever-growing breadth of diversity within *Prochlorococcus* (Table 1).

What do we know about the physiological and ecological distinctions among these clades? HLI and HLII clades are distinguished by their temperature optima[30,33] (Fig. 1C), which suggests that temperature-dependent adaptations probably had a role in their divergence. Clades

HLIII – HLVI[47-51] lack cultured representatives and are termed "HL" solely due to their phylogenetic grouping with the HLI and HLII clades. Based on their distinctive distributions along ocean transects as well as genomic and metagenomic data, it has been proposed that members of the HLIII, HLIV and HLV clades thrive in regions characterized by high N and P but low Fe availability[47-50]. There is evidence that the HLIII and HLIV clades have adapted to these iron-limited environments by decreasing cellular iron requirements[49] (Box 2) and acquiring siderophore transporters for efficient scavenging of this element[48]. Data on the HLVI clade are limited, but given that they are detected in the middle to lower euphotic zone, it has been proposed[47] that they might be adapted to lower light levels than the HLI and HLII clades.

Environmental factors associated with the diversification of LL-adapted clades are less well understood, mostly because fewer cultured representatives exist[36,39,42,52]. However, inferences can be made from ecological, physiological and genomic data. For example, the LLI clade has characteristics that are intermediate between HL-adapted and the other LL-adapted clades: they are more abundant closer to the surface and during deep mixing events in the wild[30,35] than other LL-adapted cells[29,52], and they can better tolerate fluctuating light intensities. They are also the only LL clade known to encode photolyase (a photoprotective enzyme[53]) and have more high light inducible (*hli*) genes (which encode proteins that protect cells during light shock and other stresses[54]) than any other *Prochlorococcus* clade[44,53].

In contrast to the LLI clade, the LLII/III and LLIV clades are more restricted to the lower euphotic zone and decrease in relative abundance during deep mixing events. Among cultured *Prochlorococcus*, members of the LLIV clade are the most closely related to *Synechococcus* and have the largest and most diverse genomes. They are often physically larger than HL-adapted cells and seem to produce a wide diversity of secondary metabolites[55]. The LLV and LLVI clades lack cultured representatives and have only been detected in oxygen minimum zones, where the oxygen-depleted layer meets the euphotic zone. This suggests that they are adapted to the unique redox conditions, and associated microbial community, of this habitat[47,56].

Diversity at the genomic level

The analysis of whole genomes has greatly expanded our understanding of the vast amount of genomic variation within each of the deeply branching HL and LL-adapted clades, which is evident across many levels, ranging from genome size to gene content to fine-scale allelic variation. Yet, there are clear patterns in how this diversity is organized.

Characteristics of the Prochlorococcus *genome*

Prochlorococcus is a prime example of an organism with a 'streamlined' genome[57]. Their genomes are relatively smaller than those of other cyanobacteria, which reflects a rapid decrease in size following divergence from a common ancestor with *Synechococcus*[58]. Initially, this reduction was probably driven by strong genome-wide selection for the removal of genes with only a small functional benefit that no longer justified their carrying costs[59]. Following the initial streamlining, genome diversity that correlated with the deeply branching HL and LL-adapted physiologies began to emerge. The genomes of HL-adapted strains are generally smaller in size and have a lower GC content than the genomes of LL-adapted strains (Fig 2B). Variation in these basic characteristics is also observed among LL-adapted genomes, with the LLIV clade having the largest (2.4 – 2.7 Mbp) and most GC-rich (~50%) genomes[7,44,60].

Extensive genome diversity is also apparent at the level of gene content. Although each individual isolate contains only a few thousand genes, the *Prochlorococcus* genus has a huge

pan-genome[44,60,61]. All *Prochlorococcus* isolates sequenced to date share ~1000 genes (the 'core' genome), which comprise about half of the average *Prochlorococcus* genome, and often encode basic housekeeping functions[44]. The remaining genes, which comprise the 'flexible' genome, are found in only one or a few *Prochlorococcus* genomes and presumably contribute to the relative fitness of each distinct lineage within its local environment[44,62].

The *Prochlorococcus* genome can be understood at least in part through this lens of core and flexible gene content. In contrast to other cyanobacteria (such as *Microcystis aeruginosa*), in which repeat sequences are common and genes are added and lost at a similar rate throughout the genome[63], the flexible genes of *Prochlorococcus* tend to be clustered in hypervariable islands of the chromosome[44,62]. Such genomic islands have been observed in the metagenomes of wild *Prochlorococcus* populations[62,64], and are also found in *Synechococcus*[65,66]. Although gene loss has clearly played an important part in genome evolution, gene gains have also occurred in all *Prochlorococcus* lineages; this is particularly evident in the LLIV clade[44] (see below). Genes gained by horizontal transfer (HGT) commonly occur in genomic islands, as deduced from gene occurrence patterns, homology to genes from other microbes and GC content.

The clustering of genomic hypervariability into genomic islands contributes to the maintenance of gene order in the core genome. By examining available *Prochlorococcus* genome sequences[44,60], it is observed that 45% of core genes are locally syntenic (meaning that the same genes are located immediately upstream and downstream in all isolates), compared to 30% in *Synechococcus*. In addition, HL-adapted genomes have maintained core genome synteny to a greater extent than LL-adapted genomes. Although the loss of paralogous genes is more common in *Prochlorococcus* than in *Synechococcus*, selection pressure to remove duplicate gene family members seems to be lower among the larger LL-adapted genomes than the HL-adapted genomes[67].

The Prochlorococcus *pan-genome*

The sequencing of each new *Prochlorococcus* genome adds, on average, 160 novel genes (~5-8% of the genome) to the *Prochlorococcus* pan-genome[44,60]. But what is the total number of different genes distributed throughout the global *Prochlorococcus* population? Genomes and metagenomic data show that at least 12 major clades exist (Fig. 2A) which contain over 13,000 genes (Fig. 2C). These genes are thought to have important roles in tailoring *Prochlorococcus* physiology to its local environment. Consistent with this hypothesis, isolates that occupy similar habitats, irrespective of HL or LL status, frequently have similar sets of flexible genes; such genes are often associated with nutritional adaptations (Box 2). This indicates that the flexible genes are maintained under selection, and are thus likely to contribute to fitness in these environments[68-71]. Indeed, mutations in some flexible genes have been shown to decrease growth rate, thereby reducing overall fitness[72].

The *Prochlorococcus* pan-genome continues to grow as more strains are sequenced; in addition, it is dynamic, as genes are continually gained and lost from lineages. Theoretical projections based on thirteen published genomes predicted that the global *Prochlorococcus* population contains 57,792 genes and 18 distinct clades[61]. Applying an updated version of this model[73] to 41 cultured *Prochlorococcus* genomes[44,60] increases the pan-genome size estimate to 84,872 genes – four times the size of the human genome. This analysis suggests that we have identified only a small fraction of the genes in the *Prochlorococcus* pan-genome to date.

Understanding the distribution of diversity among known clades can guide the search for 'missing' genes. For example, the genomes of LL-adapted strains contain, on average, more

unique genes than HL-adapted strains (Fig 2D). More specifically, of the genes found in any pair of LL-adapted strains, approximately 30% will be unique to each genome (measured as in[74]), whereas for pairs of HL-adapted strains only 13% of genes are unique. There is a correlation between ITS similarity and gene content similarity among *Prochlorococcus* genomes; however, LLIV strains have disproportionately more unique genes per genome than any other clade of cultured *Prochlorococcus* (Fig. 2D). Although the LL-adapted genomes are also the largest (Fig 2B), these trends are independent of genome size. These data suggest that our knowledge of the *Prochlorococcus* pan-genome will be most rapidly expanded through single-cell genomics, metagenomics and targeted strain isolation of LL-adapted cells.

But what is the cause of the increased gene content diversity among LL-adapted strains compared to HL-adapted strains? One possibility is that LL strains can acquire new genes via HGT at a higher rate than HL strains. Alternatively, it is also possible that this difference reflects the selection pressures of stable and strong environmental gradients in deeper waters – the primary habitat of LL-adapted strains – which creates additional potential niche space to select for a greater diversity of novel functions. By contrast, a significant fraction of HL-adapted cells can spend much of their lives in the more homogeneous environments of the well-mixed surface waters. The large population sizes of HL-adapted strains, and their relatively high growth rates[9,14,27,75], combine to impose strong selective pressures[45] on the relatively few niche dimensions available in this habitat, driving the system towards small variations among closely related cells.

The *Prochlorococcus* pan-genome has provided numerous insights into their contribution to ocean processes, but major gaps in gene annotations limit our ability to interpret these data. Although the metabolic functions of many core, and some flexible, genes are known, nearly 75% of the genes that are currently part of the *Prochlorococcus* pan-genome are of unknown function. In terms of understanding the biogeochemical role of *Prochlorococcus*, it is helpful that the pan-genome of the more abundant HL-adapted strains is better characterized than that of the LL-adapted cells. That said, genomes of LL-adapted strains contain a higher number of novel genes and therefore have the potential to provide clues to the functional capabilities and evolutionary history of *Prochlorococcus.* Although less abundant than their HL-adapted relatives, it seems evident that these populations have important ecological roles to play.

Fine-scale variation

In addition to differences in gene content, there is a layer of fine-scale sequence diversity that results in extensive allelic variation among bacteria. Even putatively 'clonal' *Prochlorococcus* strains can have hundreds of stably selected single nucleotide polymorphisms[45,60]. This raises the question of where the baseline of ecologically meaningful diversity lies; for example, what is the cell-to-cell diversity in a single water sample, and how does this change in response to environmental variability? Single-cell genomic analyses have recently shown that *Prochlorococcus* populations in the same milliliter of water comprise hundreds of distinct coexisting and stably maintained subpopulations[45]. Each subpopulation is associated with a unique 'genomic backbone' (a set of shared core alleles that is linked to a defined subset of flexible genes) that seems to be shaped by selection. Such backbones contain alleles that define deeply rooted adaptations as well as genes that contribute to local environmental adaptations. Even when comparing cells that have identical ITS sequences, extensive allelic and gene content diversity is observed, which seems to contribute to ecological differentiation. Population structure, as defined by genomic backbone composition, can vary over seasonal timescales but

seems to reflect ancient and stable niche partitioning, implying that this structuring of microdiversity contributes to the resilience of *Prochlorococcus*[45].

What are the mechanisms that have generated and shaped the observed variation? Although *Prochlorococcus* cells are exposed to potentially high amounts of ultraviolet radiation and lack several key DNA repair enzymes[76], the mutation rate of *Prochlorococcus* is similar to that of *E. coli* (on the order of 10^{-7} mutations per gene per generation)[77]. Thus, sequence diversity is not simply due to a high mutation rate and probably reflects the impact of the selective pressures that are imposed by the many different environments, at both the microscale and macroscale, that this genus is exposed to. From a population genetics perspective, *Prochlorococcus* has a massive effective population size that is estimated to be between 10^{11} and 10^{13} cells[45,61], which is at least four orders of magnitude larger than that estimated for *E.* coli[61] and is probably among the largest on the planet[45]. Despite the small genome size and typical bacterial mutation rate, the population size alone should minimize the impact of genetic drift and provide extensive genetic variation for selection to act on[45,59,78], thus leading to selection for minute fitness differences between strains[45].

The remarkable amount of stably maintained, co-existing genomic diversity in *Prochlorococcus* populations cannot be easily explained by classic ecotype models, in which adaptive mutations are predicted to lead to whole-genome selective sweeps resulting in a homogenous genomic population structure[79]. Selective pressures from predators (particularly phages; see below) likely play an important part in maintaining this diversity, as predicted by models incorporating density dependent fluctuating selection such as the 'kill-the-winner' and the 'constant diversity' hypotheses[80]. These are based on the idea that as a microbial lineage increases in abundance, so will the predation pressures that affect it. Thus, predation would have a disproportionately larger impact on the dominant lineage, ultimately resulting in a population of diverse genotypes with different susceptibilities to the predator. This model appears consistent with the observation of diverse genomic backbone lineages within *Prochlorococcus*; in some instances, genomic backbones link alleles of genes known to impact predation together with those affecting other physiological adaptations[45,72]. Yet, not all genes impacting predation are necessarily associated with a backbone lineage. Many such alleles are found in hyperdiverse genomic islands[72], and may recombine at a relatively high rate; if so, they would become unlinked from the backbone, and selection against these alleles would not explain variation across the genome[81]. Thus, it is likely that predator mediated fluctuating selection explains only part of the story. The maintenance of diversity within *Prochlorococcus* populations must depend on the complex and still poorly understood interplay of many forces including predation, recombination, selection, population structure, and environmental complexity.

A 'federation' of diverse cells

Prochlorococcus can be viewed as a 'federation' of co-existing cells – a large collection of many groups, each exhibiting different adaptations to specific environmental variables and representing combinatorial arrangements of alleles that reflect important niche dimensions. In turn, each of these groups contains subgroups with adaptations to slightly different selective pressures, and so on, ultimately filling out the total niche space occupied by *Prochlorococcus* (Fig. 3A). The immense diversity of *Prochlorococcus*, and particularly the combinatorial nature of this diversity, results in niche partitioning and the robustness of populations across time and

space. Different environmental conditions will select for different combinations of adaptive alleles encoded by some members of the federation, thus changing the relative abundance of different subpopulations, but ensuring that the overall *Prochlorococcus* meta-population remains stable (Fig. 3B). Thus, it could be argued[82] that the entire *Prochlorococcus* federation, comprised of cells with a backbone of core and flexible genes, might behave as its own selectable unit in microbial communities.

Interactions with phages and heterotrophs

The diversity of the *Prochlorococcus* federation can only be understood in the context of the surrounding microbial community. Below, we focus on our understanding of the interactions between *Prochlorococcus* and the cyanophages and abundant heterotrophic bacteria with which it has co-evolved, and discuss how these interactions contribute to *Prochlorococcus* physiology and diversity.

Phages as a vehicle for Prochlorococcus *genome diversification*

Cyanophages that infect *Prochlorococcus* are lytic dsDNA tailed phages that belong to the T4, T7 and lambdoid groups[83-87], and are suggested to represent a notable fraction of the total viral population in some parts of the ocean[88]. Lysogenic phages have not been found in *Prochlorococcus* genomes, even though phage integrases are present (see below) and a partial phage sequence has been detected in a partial single-cell genome[48]. The apparent absence of lysogens may be related to selection for genomic streamlining, which could lead to the rapid loss of prophages from *Prochlorococcus* genomes.

Cyanophages have played an integral part in the evolution and diversification of *Prochlorococcus* genomes. Several lines of evidence suggest that phage-mediated HGT is important for gaining flexible genes in genomic islands (Fig. 4A). For example, tRNA genes, which are common sites for the integration of phages[89], often flank genomic islands in *Prochlorococcus*[62]. In addition, several genes including integrases, DNA methylases, and stress response genes are found in both genomic islands and cyanophage genomes[62]. The upregulation of several of these genes during phage infection has led to the hypothesis that host genes expressed during infection have been stably incorporated into phage genomes, which increases their opportunity for transfer back to *Prochlorococcus*[90]. A striking example of this is the expansion of the *hli* gene family of stress response genes in *Prochlorococcus,* which was most likely mediated by phages[62,90,91].

The influence of phages on gene sequence diversity is not limited to the flexible genome. Intragenic recombination between core photosynthesis genes shared by both *Prochlorococcus* and their phages seems to accelerate the diversification of the genes that encode this key metabolic process[92,93]. This could be a general phenomenon that affects those genes that are found in both *Prochlorococcus* and their phages.

Prochlorococcus and practically all marine *Synechococcus* lack CRISPR–Cas systems for phage resistance[94], perhaps because it is too costly to maintain these genes or because CRISPR–Cas systems may be ineffective given the high levels of phage diversity that is predicted for cyanophages[95]. In addition, most *Prochlorococcus* and *Synechococcus* strains also lack restriction-modification systems, which suggests that restriction is also not a widespread mechanism of defence against cyanophages in these cyanobacteria[94]. Instead, resistance to cyanophages in *Prochlorococcus* is typically conferred by mutations in genes encoding cell-surface molecules, which impair phage attachment to the cell surface[72]. These genes are part of

the flexible genome, have been laterally acquired from other bacterial phyla, are located in genomic islands and account for the greatest genomic differences between closely related *Prochlorococcus* strains[44,62,72]. This strongly suggests that selection pressure from phages, and potentially grazers (see below), influences both sequence diversification and presence of genes encoding cell surface molecules and their biosynthesis (Fig. 4B). Furthermore, phage selection presumably leads to the loss of such genes and the gain of non-orthologous genes with equivalent functionality that compensate for such losses[72].

Mutations that confer phage resistance often have a fitness cost, which manifests as either a reduction in growth rate or enhanced susceptibility to other phages[72]. For example, the mutation of cell surface molecules provide resistance to a subset of phages, but can also confer enhanced susceptibility to other phages. In this way, phages contribute to the diversity and population structure of *Prochlorococcus*: each population is composed of an assortment of subpopulations that differ in susceptibility to the diversity of phages found in the oceans[72,96] (Fig. 4B). This variability probably leads to fluctuations in the abundance of host and phage subpopulations, preventing a high degree of infection at the population level and thus facilitating stable coexistence between *Prochlorococcus* and their phages[72,80,96]. These host-phage dynamics suggest that phages may have a limited ability to control the size of *Prochlorococcus* populations but have a strong influence on population structure and diversification.

Phage influences on host metabolism

Most cyanophage genomes encode bacteria-like genes, many of which have been acquired from their cyanobacterial hosts[84,86,91,97,98]. These 'host-like' genes are often found in islands on the phage genome[86,87,98] and probably influence the infection process (Fig. 4C). For example, although phage infection disrupts *Prochlorococcus* gene expression[90], key metabolic processes such as photosynthesis are sustained, in part because cyanophages encode and express homologues of key genes in photosynthetic pathways[91,97,99]. In addition, phage-encoded Calvin cycle inhibitor genes and pentose phosphate pathway genes are transcribed together with genes involved in photosynthesis, DNA replication and metabolism[70,90]. This suggests that the phage directs the energy derived from photosynthesis away from carbon fixation and towards the pentose phosphate pathway to produce reducing power for deoxynucleotide biosynthesis. Indeed, the ratio of NADPH to NADP is increased in phage-infected *Prochlorococcus* cells[70]. Similarly, host-derived phosphate acquisition genes are transcribed from phage genomes during infection, exhibit higher expression during infection of phosphate-deprived host cells and are even regulated in the phage by the phosphate two-component regulatory system of the host[100].

The role of heterotrophs

Prochlorococcus has a central role in supplying photosynthetically fixed carbon to the marine heterotrophs with which they co-exist. For example, members of the abundant SAR11 clade require glycine or serine for growth, a requirement that can also be fulfilled by their metabolic precursor glycolate[101], which *Prochlorococcus* releases in substantial amounts[102]. The heterotrophic community, in turn, influences the fitness of *Prochlorococcus*, as evidenced by the difficulty of removing heterotrophic 'contaminants' from *Prochlorococcus* cultures in the early days of its cultivation[103]. Since then, axenic strains have been generated by a variety of approaches[104-107], which enable interactions between *Prochlorococcus* and co-cultured heterotrophs to be studied in a more systematic manner. The presence of certain heterotrophic bacteria can increase the growth rate of *Prochlorococcus*, final culture density and the longevity

of the cultures, whereas other heterotrophs have inhibitory or neutral effects[107,108]. Although the mechanisms underlying the inhibitory interactions are not understood, some insights into the beneficial interactions have been revealed.

An elegant set of laboratory and field experiments has shown that *Prochlorococcus* grows better in the presence of some heterotrophs because they reduce the concentrations of toxic reactive oxygen species (ROS; such as hydrogen peroxide), which compensates for the absence of catalase and peroxiredoxin genes in *Prochlorococcus*[107,109]. This led to the 'Black Queen' hypothesis[110], which posits that free-living microbial communities evolve and sustain a division of labour for certain essential functions. In this scenario, a subset of cells carries out an essential function that becomes a 'public good', which enables non-producing cells to benefit from this activity and avoid the expense of performing it. Due to the strong selective pressure on all cells to avoid damage by ROS, cells that dispense with the costly expression of defence mechanisms have an advantage as long as they can rely on other cells for protection. In this case, *Prochlorococcus* does not produce catalase but is protected from ROS from nearby heterotrophs[110].

Prochlorococcus undoubtedly impacts heterotrophs in other ways. For example, certain strains produce a remarkable diversity of lanthipeptide secondary metabolites[55] and although their function in *Prochlorococcus* is unknown, similar compounds have functions ranging from antibiotics to surfactants[111]. In addition *Prochlorococcus* continually releases small (~100nm diameter) extracellular membrane vesicles[112] that contain a wide variety of components, including lipids, proteins, as well as small fragments of DNA and RNA. Although the ecological function of these vesicles is currently unknown, they might function as vehicles for the movement of carbon through marine food webs, as vectors for HGT or possibly decoys for predators and phage.

Distributing the genome through a community

The concept of the pan-genome is based on the notion that the total genetic repertoire of a bacterial group is greater than the number of genes encoded in any single strain[113]. Considering the Black Queen hypothesis and the impact of phage-encoded homologues of bacterial proteins on cellular physiology, should the pan-genome of *Prochlorococcus* be broadened to include heterotroph-encoded genes that supply essential functions for *Prochlorococcus* survival*?* Should it also include phage-encoded genes that function in the host? In keeping with the view that entire microbial communities are relevant units of biological organization[114], certain heterotrophs and phage could be part of the same selectable unit as *Prochlorococcus*[82], which would strengthen arguments for their inclusion in the same pan-genome. Although it is not clear where to draw boundaries to define the 'complete' metabolic repertoire of one organism, co-evolutionary selection pressures undoubtedly lead to some tight associations. Identifying discontinuities in the network of interactions could help reveal these associations and expand the *Prochlorococcus* pan-genome.

Prochlorococcus **and ocean carbon cycling**

Prochlorococcus is an important global primary producer, especially in the oligotrophic ocean where dissolved organic carbon from this group contributes up to 40% of total bacterial production[102]. *Prochlorococcus* releases a diverse range of organic molecules into the surrounding seawater[115] by a variety of mechanisms. These include direct secretion (sometimes termed 'leakage') from the cell[102,112] and cell lysis, which is mediated either by phage or by

feeding of grazers[116]. *Prochlorococcus* also directly supports the carbon and nutrient requirements of other trophic levels as it is prey to a wide range of eukaryotic predators, including tunicates[117,118], ciliates[119,120], flagellates[120,121], prymnesiophytes[122], stramenopiles, and dinoflagellates[123]. It is possible that mixotrophic eukaryotes that have primarily autotrophic lifestyles feed on *Prochlorococcus* as a source of N and P. Since acquisition of these essential elements is limited by diffusion in large cells, direct transport may be insufficient to support their nutrient requirements, whereas engulfing concentrated 'packets' of nutrients in the form of small cells like *Prochlorococcus*[123,124] may provide an advantage in hyper-oligotrophic environments.

Much of the carbon fixed by *Prochlorococcus* in the euphotic zone is believed to be recycled in the upper waters through the microbial loop: it is taken up by heterotrophic bacteria and is either respired or incorporated into other compounds that move up the food web[116]. However, it has also been reported that *Prochlorococcus*-derived carbon is exported to deep waters by aggregation and sinking of biomass following trophic processing[125]. For example, degradation products of the unique *Prochlorococcus* divinyl chlorophyll *a* have been found in the faecal matter of salps recovered from deep waters[117]. The degree to which *Prochlorococcus* participates in this biological pumping of carbon from the atmosphere to the deep ocean is an open question.

Our understanding of the role of *Prochlorococcus* in marine carbon cycles has been complicated by a growing body of evidence that mixotrophy – the use of both photoautotrophic and heterotrophic modes of growth – occurs in both cultured and wild *Prochlorococcus* populations. *Prochlorococcus* can import organic compounds for use as either N, P, energy or carbon sources. High uptake rates of amino acids, including both methionine and leucine, have been observed in wild *Prochlorococcus* populations[126-129], which are capable of assimilating nucleic acids, possibly functioning as a nitrogen source[127]. In addition, studies of both cultured and wild *Prochlorococcus* have shown that they can take up glucose[130,131]. This is particularly intriguing because glucose lacks both nitrogen and phosphorus; thus, it could only be used as a source of carbon or energy.

Prochlorococcus in a warming world

Advancing our understanding of the ecology and physiology of *Prochlorococcus* is particularly important in the face of global climate change. The rise in surface water temperatures and the expansion of ocean stratification will almost certainly impact the structure and function of the bacterial populations. For example, models predict that in a world with ~650 ppm atmospheric CO_2, the global abundance of *Prochlorococcus* may increase by over 25% and expand their range towards the poles as the waters increase in temperature[3]. The complexity of these scenarios in terms of the distribution of ecotypes is daunting, but hypothetical scenarios can be illuminating. An expansion of stratified waters will decrease nutrient input from the deep, making these regions more oligotrophic, which will almost certainly change the local ecotype distributions. We expect that members of the HLII clade, which are currently the most abundant group and have the highest optimum temperature for growth, will expand their habitat into higher latitudes. By contrast, the relative abundance of groups with lower temperature optima would shift away from the equator.

As we consider such scenarios, it is important to recognize that selection for genome-wide adaptations, such as temperature optima, will simultaneously select for linked traits that are encoded by the same genomic backbone, such as specific nutrient assimilation capabilities. Because *Prochlorococcus* biomass is a significant fraction of total photosynthetic biomass, their

biogeochemical contributions would then feedback on the environment and shift selection pressures in the entire ecosystem. Thus *Prochlorococcus* subpopulations, with different physiological abilities, will arrive in different regions of the oceans through a complex set of feedback loops. Although climate change may increase the abundance of *Prochlorococcus* worldwide, we cannot predict how the complicated interrelationships between the cell, its community and the environment will eventually play out; the system is simply too complex.

Future Challenges

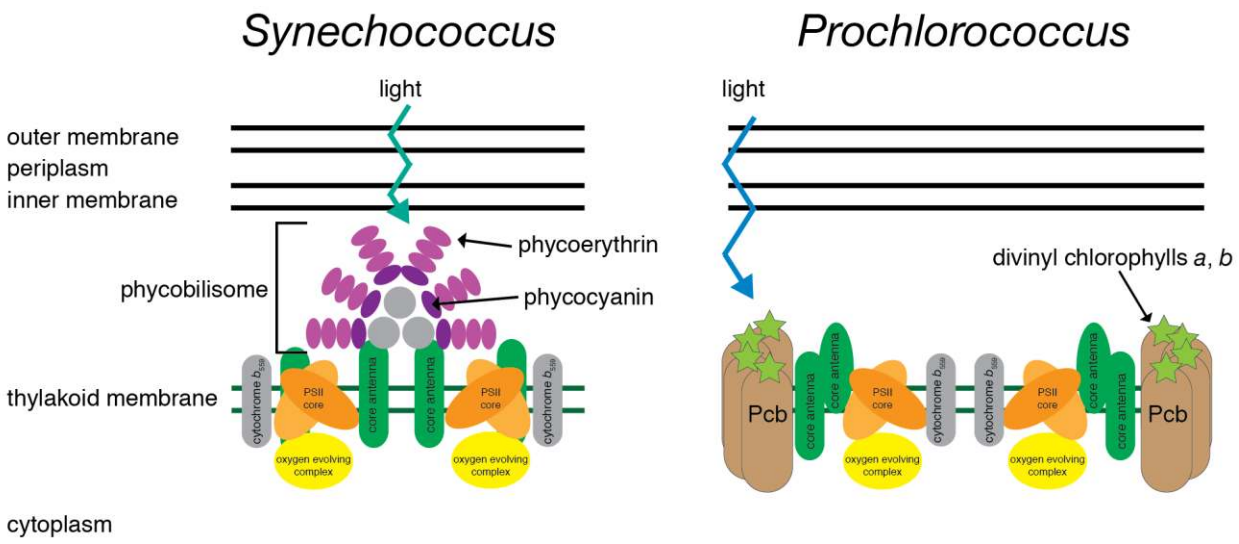
The pace of discovery in studies of *Prochlorococcus* ecology and evolution increased dramatically after the first genomes became available and continues to increase as a result of improvements in the technologies available for DNA sequencing. In the next decade we should get closer to describing the global pan-genome of *Prochlorococcus* and the distribution of its genes among different regions and along vast oceanic gradients. Interpreting these data in an integrated physiological and ecological context represents an enormous challenge, which will ultimately require unraveling the function of the large number of unannotated genes in microbial genomes. Genes of unknown function hold important clues about how microbial ecosystems function and deciphering the metabolic function of those that are unique to *Prochlorococcus* will be particularly illuminating for understanding its role in the ecosystem. Advances in this area will require the development of an efficient genetic system for *Prochlorococcus,* which has proven to be a challenge.

Full exploitation of the information from metagenomic, metatranscriptomic and single cell genomic data from field studies relies heavily on reference genomes and physiological studies of cultured strains. Continued efforts to obtain new isolates of *Prochlorococcus* from diverse regions of the ocean – along with the abundant oligotrophic, heterotrophic bacteria and phage with which it co-exists – will be important in this regard. These cultures are essential for testing hypotheses about the forces that shape these co-evolved genomes and the global biogeochemical influence of *Prochlorococcus* and its metabolic partners. Finally, although we know by inference that the death rates of *Prochlorococcus* are sizable in the wild, our understanding of the impact of viral infection, predation, and spontaneous cell death in stabilizing these populations is in its infancy. There is much yet to learn!

Acknowledgments

The authors would like to thank members of the Chisholm lab for providing helpful comments on the manuscript. We also thank J. Waldbauer for performing the initial calculations which inspired Fig. 1B. SB, PB and SWC were supported by grants from the Gordon and Betty Moore Foundation (Grant GBMF495 to SWC) and the National Science Foundation (OCE-1153588, OCE-1356460, and DBI-0424599, the NSF Center for Microbial Oceanography Research and Education). DL was supported by the Israel Science Foundation (Morasha Grant 1504/06) and the European Research Council (Starting Grant 203406).

Box 1: *Prochlorococcus* and *Synechococcus*: what's in a name?



Prochlorococcus and marine *Synechococcus* are thought to have diverged from a common ancestor ~150 million years ago[76], but on the basis of the 16S rRNA sequence, most members of the two groups would be considered the same species. Although differences in cell size and photosynthetic pigments yield distinct flow cytometry profiles, they still share many phenotypic and ecological traits. Whole-genome phylogenies clearly separate *Prochlorococcus* from *Synechococcus*[44], but the phylogenies of many individual gene families cluster low-light (LL) adapted *Prochlorococcus* more closely with *Synechococcus* than with high-light (HL) adapted *Prochlorococcus*[37,132]. While these data question the distinction between these two genera, several physiological and ecological factors justify their separation.

The clearest difference between the two groups is in their photosynthetic apparatus (see the figure). Similarly to most cyanobacteria, the major light-harvesting antenna in *Synechococcus* is the phycobilisome, which is comprised of phycobiliproteins (for example, phycoerythrin or phycocyanin), each of which binds one or several light-harvesting chromophores such as phycoerythrobilin or phycocyanobilin[133]. This antenna complex collects light and transfers the energy to the photosystem II (PSII) core antenna proteins (CP43 and CP47) and then into the PSII reaction center (comprised of multiple proteins and cytochrome b_{559}). *Prochlorococcus* is one of the few cyanobacteria (together with *Prochloron* spp. and *Prochlorothrix* spp.) that lack phycobilisomes; instead, its main light-harvesting antenna complex is made up of proteins (the prochlorophyte chlorophyll-binding protein, Pcb) that bind divinyl chlorophyll *a* and divinyl chlorophyll *b*. *Prochlorococcus* also uses monovinyl chlorophyll *b* as an accessory pigment in the antenna complex[9,58,133]; together, these unique pigments enhances their absorption of blue light[1,8], which are the dominant wavelengths in deep waters.

The geographic distributions of *Prochlorococcus* and *Synechococcus* provide clues about the forces that mediate their niche partitioning. *Synechococcus* is present in almost all marine environments, whereas *Prochlorococcus* is restricted to warmer, oligotrophic oceans such as subtropical gyres and the eastern Mediterranean Sea, and is absent from colder, nutrient-rich waters at high latitude and in most nutrient-rich coastal waters. What might explain these differences? *Synechococcus* can tune its phycobilisome antenna systems to acclimate to changing

temperatures, which may contribute to its greater geographical range[134,135]. It is also less susceptible to copper toxicity than *Prochlorococcus*[136] which might explain, in part, its dominance in coastal waters. Furthermore, *Synechococcus* strains have higher maximum growth rates[9] than *Prochlorococcus* and they are prey for many of the same predators. Because the growth rate of predators is coupled to that of their prey[137] it may be impossible for *Prochlorococcus* to achieve net positive growth rates when *Synechococcus* is growing maximally; it would simply be 'grazed away'[137]. Consistent with this hypothesis, *Prochlorococcus* can be cultivated in the lab using seawater collected from coastal sites[138] even though it is essentially absent from such regions in the wild.

Figure modified from Ref[133].

Box 2: Nutrient acquisition genes as signatures of the local environment

Prochlorococcus strains vary in their ability to utilize different inorganic nutrient sources, and much of this physiological diversity is clearly reflected in their underlying genomic diversity[58]. Adaptations linked to the availability of P, N and trace metals do not follow the ribotype-defined phylogeny as observed for light and temperature[36], and are better interpreted as signatures of the local environment in which a given strain is found[69]. Thus much can be learned about the environment and selective forces experienced by a given *Prochlorococcus* cell through the composition of its genome, and in some cases through the composition of the cell itself.

Phosphorous

Genomic and metagenomic analyses have revealed that *Prochlorococcus* populations in phosphorus-limited environments, as well as the cyanophage that infect them, contain relatively more genes involved in phosphorus acquisition compared to populations from environments where phosphorus is more abundant[68,69,71,139,140]. *Prochlorococcus* genes involved in phosphite and phosphonate assimilation are also prevalent specifically in *Prochlorococcus* populations from phosphorus-limited environments[140-142]. P-starvation experiments in the laboratory have revealed that in addition to known P-starvation response genes, several genes of unknown function, all of which are clustered in a hypervariable genomic island, are highly up-regulated[62,68]. Unraveling the functions of these genes will shed light on the response of these cells to phosphorous stress.

Nitrogen

Productivity in many regions of the oligotrophic ocean is limited by nitrogen availability; indeed, the average amount of N in the *Prochlorococcus* proteome (per amino acid sequence) is reduced relative to that of coastal bacteria[143]. N minimization is attributable, in part, to the low GC composition of *Prochlorococcus* genomes: the amino acids encoded by low GC codons have a lower N content (reduced N:C ratio) than those encoded by GC-rich codons[144]. Surface waters tend to be more N-limited than deeper waters; this correlates with the fact that HL-adapted strains, which are typically most abundant near the surface, have a lower GC content – and thus require less N – than LL-adapted strains[145]. Some strains exhibit additional signatures of selection for N minimization in the particularly reduced N content of many N stress-responsive proteins[145].

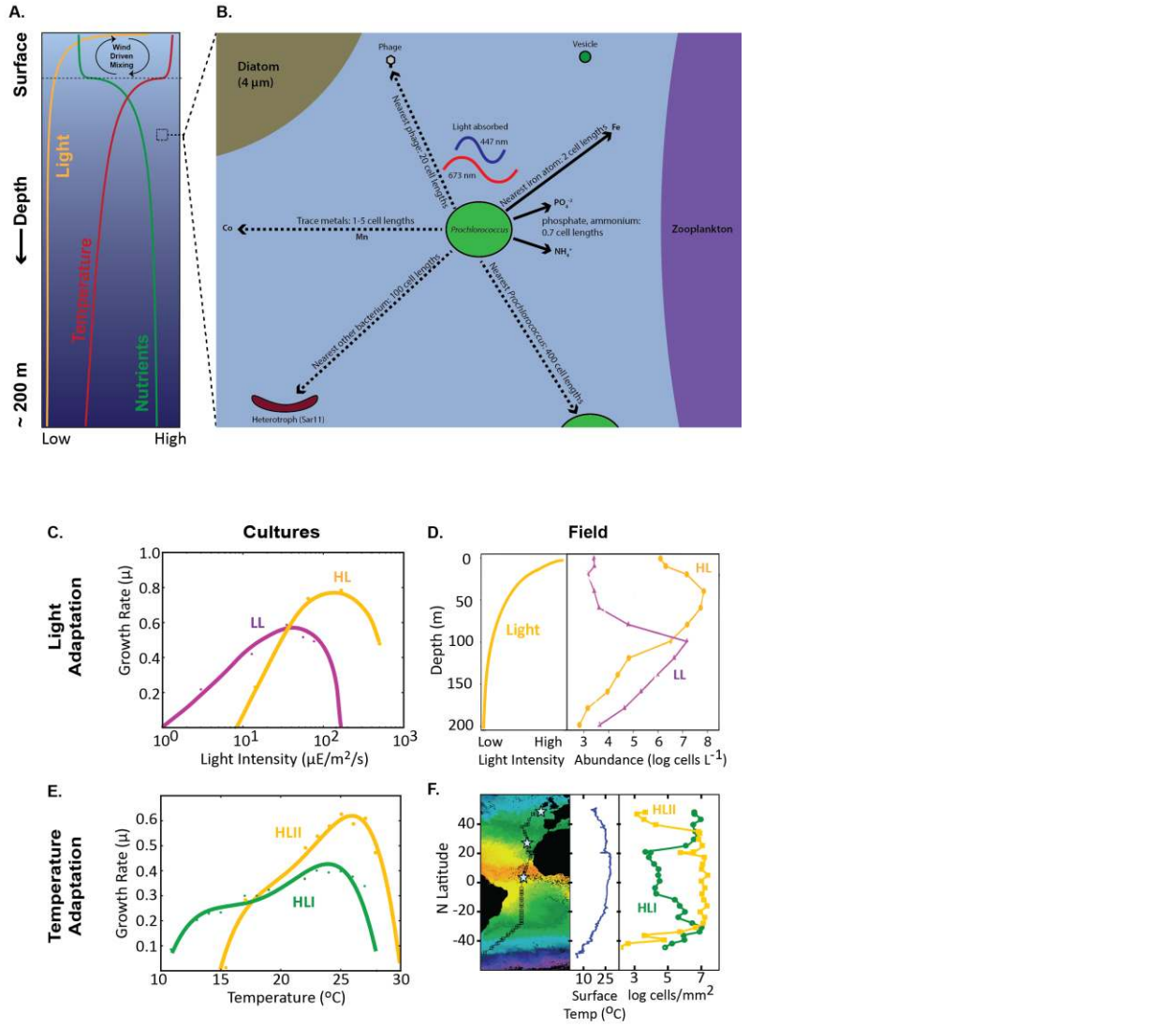
Although all *Prochlorococcus* strains can use ammonium, and none can fix dinitrogen, they differ in their ability to assimilate other forms of N including urea, cyanate, nitrite, nitrate and amino acids[53,106,126,146-149]. Nitrite, cyanate and amino acid uptake genes seem to be subject to horizontal gene transfer (HGT), as indicated by their positioning in genomic islands of some strains[7,62]. Genomic analysis of recently identified nitrite and nitrate assimilation genes suggests that nitrate assimilation may have been maintained in distinct lineages of the LLI and HLII clades for some time[106,147]. Nevertheless, genes associated with nitrate assimilation also seem to be subject to HGT, as suggested by the discovery of common mobility elements surrounding the nitrate assimilation genes in one genome[106].

Iron

Because of its importance in photosynthetic reaction centres and its low concentration in ocean waters, iron availability seems to exert substantial selection pressure on *Prochlorococcus* niche

differentiation. Cultured strains show large variations in their iron requirements; for example, the LLIV strain MIT9313 can grow at an iron concentration that is an order of magnitude lower than that required by the HLI strain MED4[150]. Cells from the uncultured HLIII and HLIV clades, which have been found in iron-limited regions, may have reduced their iron requirements by dispensing with several iron-containing proteins including cytochrome C_m, two ferrodoxins, and the plastoquinol terminal oxidase[49]. There is also evidence to suggest that these cells may use siderophores to increase iron acquisition[48]. The diversity in iron acquisition and the requirement for iron among different *Prochlorococcus* strains is consistent with the observation that many genes that show differential expression during iron starvation have signatures of HGT[150].

Figure 1: The world of *Prochlorococcus*



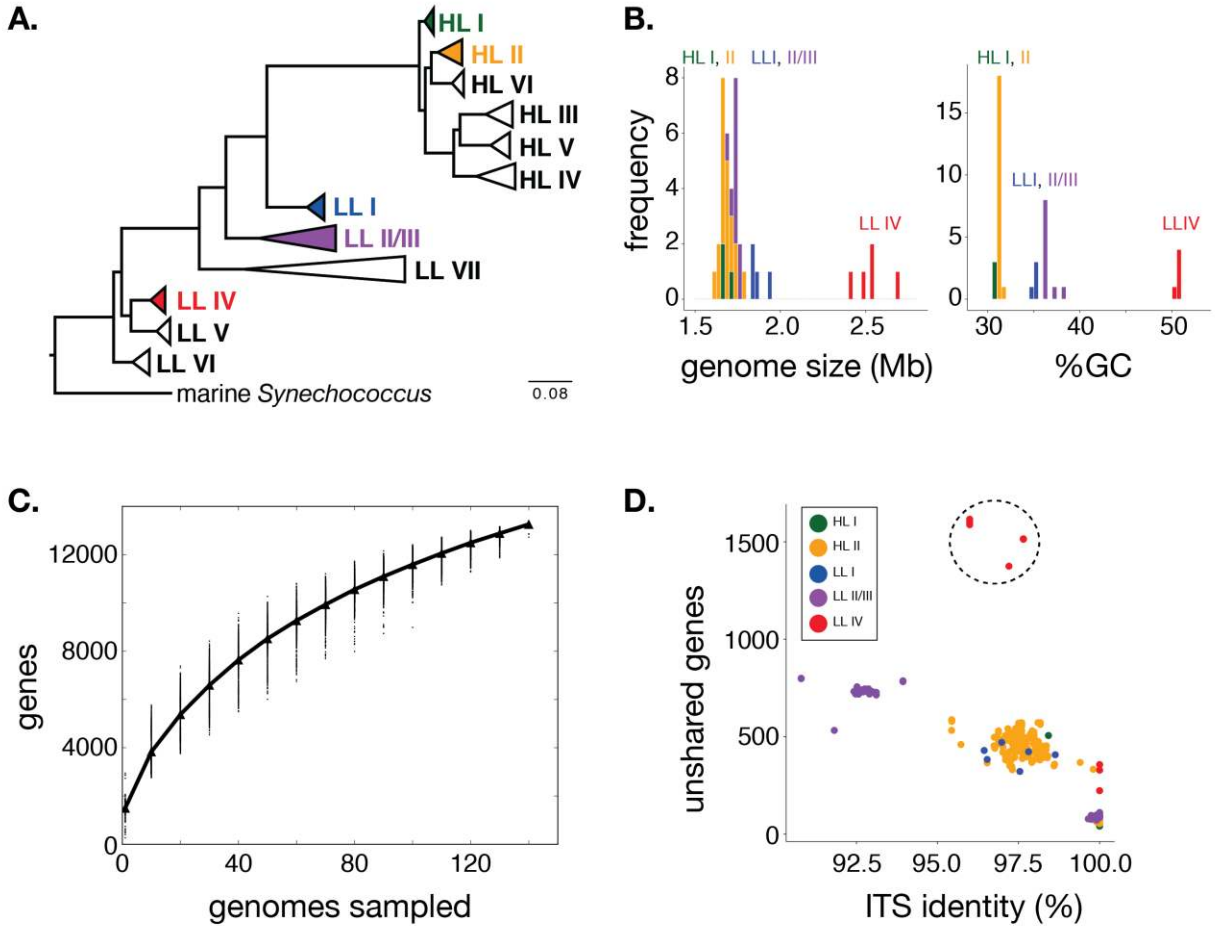
(A) *Prochlorococcus* inhabits the entire euphotic zone, which is characterized by gradients of light, temperature and nutrients. The ocean water column is divided between an upper 'mixed' layer (where wind and heat-driven turbulence homogenizes the distribution of nutrients and cells) and the deeper, stratified, and less turbulent waters where gradients in nutrients form as a result of biogeochemical activity. Light levels decrease exponentially with depth; gradients of temperature and nutrient concentrations are largely similar in the mixed layer, while temperature decreases and nutrient concentrations increase with depth.

(B) The oligotrophic ocean represents an extremely dilute environment in terms of organisms and nutrients. Some key players in oligotrophic marine communities are shown, along with the distances between them as estimated from their average concentrations. This leads to a simplified conceptual framework for considering the context in which *Prochlorococcus* evolved. According to these metrics, each *Prochlorococcus* cell should be hundreds of cell diameters away from other members of its 'federation' and even a few cell diameters away from essential nutrients. The closest phage (of any kind, let alone one that might infect *Prochlorococcus*) might be tens of

cell lengths away. Mean inter-particle distances between a component of the ocean and *Prochlorococcus* are based on average concentrations in the North Pacific[35]. Distances that are to scale are marked with a solid line, whereas those that are not to scale are represented by dashed lines. Light wavelengths represent the two major absorption peaks of *Prochlorococcus* MED4[133].

(C-F) The relationship between growth optima of cultured isolates in the lab and their abundance in the field has been most clearly demonstrated with both longitudinal and depth niche partitioning of ecotypes as a function of light (C,D) and temperature (E,F). The relative abundance of HL- and LL-adapted strains in the water column is consistent with light optima in the lab, and the longitudinal abundance of two HL-adapted clades are consistent with the temperature optima for growth of representative strains, in which cells from the HLII clade grow maximally at higher temperatures than HLI strains. Although growth rate and abundance need not be related in wild populations, so far these patterns are correlated without exception, making it easier to interpret distribution patterns in the wild in the context of the physiology of the organism. Figure 1C adapted from[27,33].

Figure 2: *Prochlorococcus* phylogenetic and genomic diversity



(A) *Prochlorococcus* phylogeny as seen through rRNA ITS sequence diversity. Twelve major clades are known, five of which (in color) have cultured representatives; the rest are only known through environmental sequence data. The division between HL and LL clades correlates with adaptations to different light optima that mirror ribosomal phylogeny.

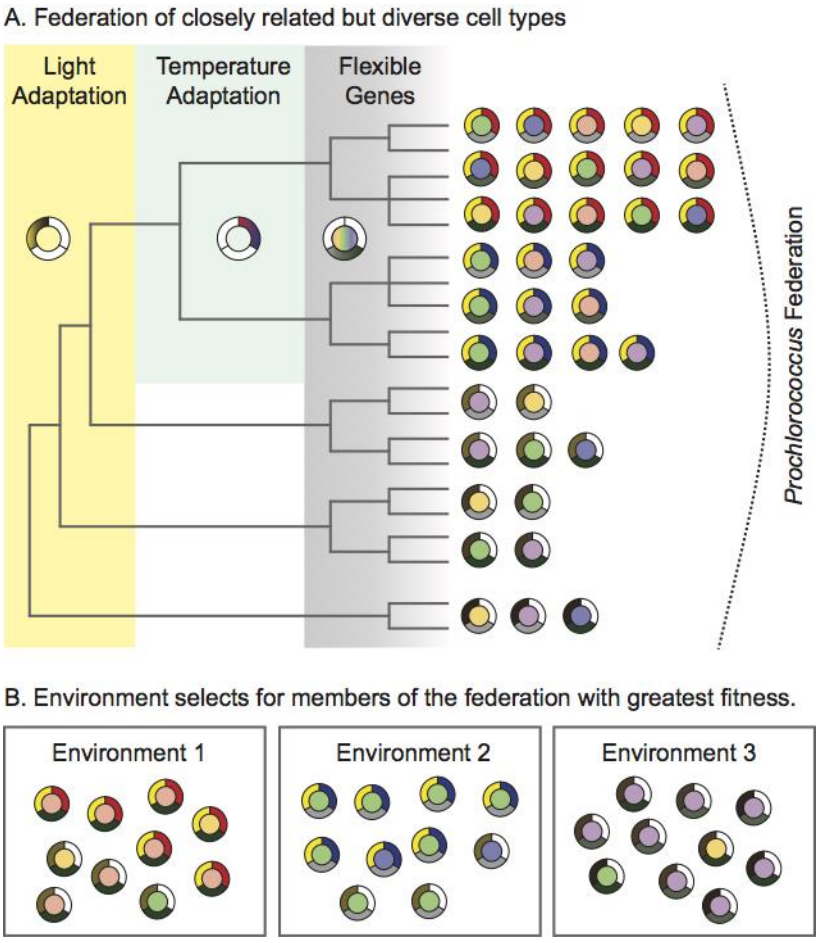
(B) Distributions of genome size (left) and %GC content (right) of cultured *Prochlorococcus* isolates demonstrates that these characteristics correlate with phylogeny. Genomes of HL-adapted cells have, on average, the smallest and least GC-rich genomes among *Prochlorococcus* and most other cyanobacteria. Genomes from members of the LLIV clade are the largest and most GC-rich, and genomes from the LLI and LLII/III clades falling between these two extremes.

(C) Analysis of the pan-genome shows that the number of novel genes in the *Prochlorococcus* federation increases with each additional genome (calculated as in[44]). Data include cultured isolates, single-cell genomes and consensus metagenomic assemblies.

(D) Gene content diversity varies among *Prochlorococcus* clades. Compared to the HL-adapted strains, LL-adapted strains contribute more new genes to the *Prochlorococcus* pan-genome. Each

point represents a pairwise comparison of two genomes in each clade (indicated by color). Overall, the number of unique genes found in any pair of genomes is inversely correlated with ITS identity (although strains with identical ITS sequences still differ in gene content). LLIV genomes (red; highlighted by the broken circle) are clearly unique among *Prochlorococcus*: they have the greatest average gene diversity of any clade, which cannot be explained by their ITS diversity.

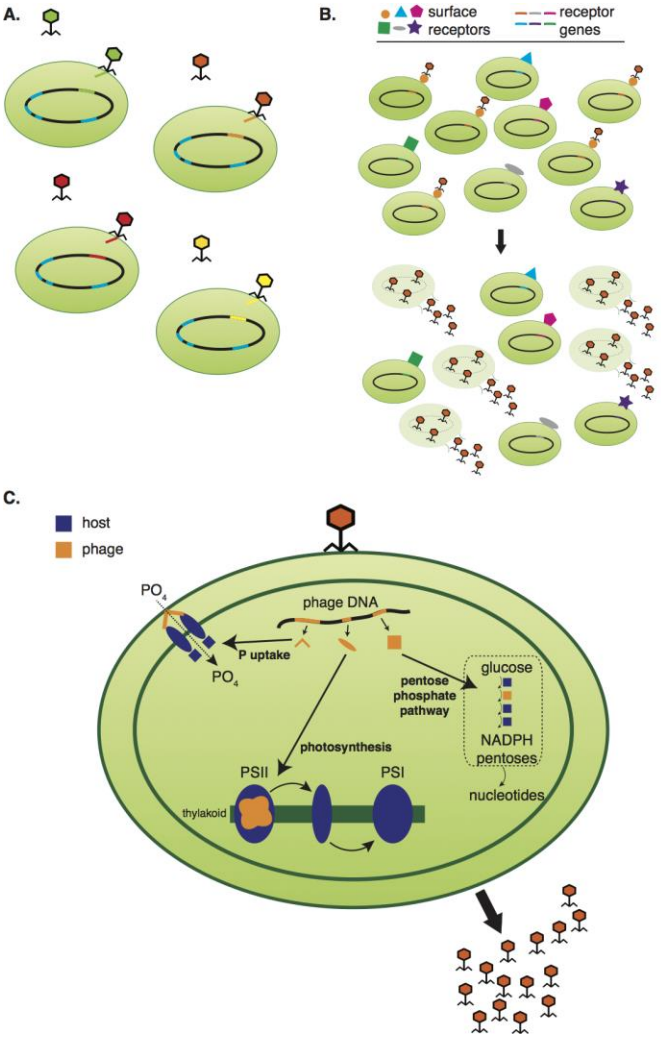
Figure 3. The *Prochlorococcus* "federation"



(A) The *Prochlorococcus* "federation" is composed of groups of cells, each representing different combinatorial arrangements of genes that are required for adaptation to their distinct ecological niches. Each circle represents an individual *Prochlorococcus* cell or clonal lineage. The outer coloured ring represents the genomic 'backbone' of the cell (which contains both core and flexible genes) and the inner colour represents a unique set of flexible gene content. The genomic backbone consists of alleles that determine adaptation to basic, deeply divergent traits such as light and temperature optima for growth, together with a subset of flexible genes that contribute to niche adaptation[45]. The composition of the backbone generally correlates with the whole-genome phylogeny but the flexible gene content varies markedly according to the local environment.

(B) The diversity found within the federation contributes to the stability and resilience of global *Prochlorococcus* populations by providing an extensive pool of diversity that different environmental conditions can select.

Figure 4: The influence of phages on *Prochlorococcus* gene content, population diversity and host physiology



(A) *Prochlorococcus* cells have a common core genome (in black) that encodes basic housekeeping functions, in addition to a diverse set of flexible genes that are primarily located on genomic islands (other colours of the genome), that seem to be under strong selection. Phages alter the genetic content of these islands by transferring genetic material (which may have been acquired from another cell; indicated by coloured DNA injected by the phage) from the phage into the infected cell.

(B) To infect and subsequently lyse a cell, phages attach to cell-surface molecules, which results in the selection for cells with surface molecules other than those recognized by abundant phages in the surrounding environment. Thus, predation by phage can prevent any one genotype with a particular set of cell surface molecules from becoming dominant within a population, fostering population diversity. Cell surface genes are often encoded in genomic islands. Selection for diverse cell-surface genes together with horizontal transfer of genes into genomic islands lead to

a large *Prochlorococcus* population with a common core genome but an assortment of subpopulations with different flexible genomes[72,80].

(C) Cyanophages encode bacterial-like proteins that can participate in cellular metabolic processes together with host proteins. These host-phage hybrid complexes presumably facilitate the acquisition of energy and materials needed for DNA synthesis and the production of high numbers of phage progeny. The host processes that seem to be temporarily boosted by phage genes (based on gene expression analysis) include photosynthesis for energy production, transport of phosphate for nucleotide biosynthesis, and pentose phosphate pathway (PPP) proteins for reducing power and nucleotide precursors. Proteins derived from the host genome are shown in blue and phage-encoded proteins are shown in orange.

Table 1. The major clades of *Prochlorococcus* as defined by rRNA ITS sequences.

Clade Name	Alternate names in the literature for the same ribotype	Representative cultured strains[c]	Habitat where relatively more abundant, and/or where isolated
HLI	eMED4[29], Low-B/A *Prochlorococcus* clade I[37]	MED4, MIT9515	Isolated from the upper/mid euphotic zone, typically from the subtropical ocean. Their distribution is shifted to higher latitudes, due to a relatively low temperature growth optimum[31,33-35].
HLII	eMIT9312[29], Low-B/A *Prochlorococcus* clade II[37]	AS9601, MIT9215, MIT9312, SB	Often found throughout the euphotic zone; typically among the most abundant *Prochlorococcus* group in the water column. Especially abundant at lower latitudes, due to a relatively high temperature growth optimum[31,33-35].
HLIII	HNLC1[47,50]; HNLC2[49,b]	None	Sequences from high nutrient, but low chlorophyll containing equatorial waters, typically between 10 °N – 10 °S in the Pacific and Indian oceans. These regions are typically limited by iron availability, and sequence data suggests that these cells have adaptations for reducing cellular iron requirements[47-50].
HLIV	HNLC1[49]; HNLC2[47,50,b]	None	Sequences from high nutrient, but low chlorophyll containing equatorial waters, typically between 10 °N – 10 °S in the Pacific and Indian oceans. These regions are typically limited by iron availability, and sequence data suggests that these cells have adaptations for reducing cellular iron requirements[47-50].
HLV		None	HLV sequences have been found in surface equatorial waters typically limited by iron availability. Physiological distinctions between the HLIII, HLIV and HLV clades are not known[47].
HLVI		None	Sequences were identified in the mid/lower euphotic zone (75-150m) of the South China Sea. This group has been postulated to have an intermediate light optimum[47].
LLI	eNATL2A[29], High-B/A *Prochlorococcus* clade I[37]	NATL1A, NATL2A, PAC1	Typically most abundant in the middle euphotic zone of stratified waters. Unlike other LL clades, they often remain abundant in mixed waters throughout the water column due to their ability to tolerate light shock[35,37].
LLII/III[a]	eSS120/eMIT9211[29], High-B/A *Prochlorococcus*	MIT9211, SS120	Typically found in the middle/lower euphotic zone[35,37,44].

	clade II/III[37]		
LLIV	eMIT9313[29], High-B/A *Prochlorococcus* clade IV[37]	MIT9303, MIT9313, MIT0701	Typically most abundant near the base of the euphotic zone; highly susceptible to light shock[35,37].
LLV		None	Found maximally abundant in the lower euphotic zone of oxygen minimum zones when oxygen depleted layers extend into the upper water column[56].
LLVI		None	Found maximally abundant in the lower euphotic zone of oxygen minimum zones when oxygen depleted layers extend into the upper water column[56].
LLVII	NC1[36]	None	Sequences were identified in the lower euphotic zone of subtropical waters; little is known about this clade[36].

[a] Originally defined as separate clades[37], the LLII and LLIII clades are now grouped because their separation is not well resolved phylogenetically.

[b] Two publications[49,50] assigned the names HNLC1 and HNLC2 to different clades; moving forward, we suggest the use of the HLIII and HLIV nomenclature to refer to these clades[47,48].

[c] For more information on these and other strains, see[10,44,60] and references therein.

Glossary:

Autotrophic: refers to the ability to build complex, energy-containing organic molecules from carbon dioxide using either light or inorganic chemical reactions as an energy source. *Prochlorococcus* is capable of photoautotrophic growth and uses light energy to turn CO_2 into organic carbon via photosynthesis.

Ecotype: A genetically and physiologically differentiated subgroup of a species that occupies a distinct ecological niche.

Oligotrophic: An environment with low concentrations of available nutrients.

Clade: A coherent phylogenetic group of organisms that have evolved from a common ancestor.

Planktonic: Free-floating organisms that are unable to swim against a current, making them essentially drifters.

ITS sequence: The Internal Transcribed Spacer, a non-functional rRNA sequence located between the 16S and 23S ribosomal genes in bacteria, which is a useful phylogenetic marker.

Euphotic zone: The sunlit upper region of the ocean water column that receives sufficient light energy to sustain photosynthesis. The depth can vary depending on local conditions, but it is generally the upper ~200m in oligotrophic waters.

Synteny: The conserved ordering of genes along a chromosome.

Pan-genome: The complete set of genes that is encoded by all the genomes of a defined group of organisms.

Genetic drift: the change in the frequency of an allele in a population due to chance or random events.

Effective population size: In population genetics, the size of an idealized population that would be expected to behave the same as the actual population in terms of the effects of selection and genetic drift.

Calvin cycle: The biochemical process that converts CO_2 into glucose.

Reducing power: In redox chemistry, the availability of compounds that can supply electrons.

Reactive oxygen species: Any of a number of oxygen-containing compounds, such as H_2O_2, which readily react with and damage cellular components.

Extracellular membrane vesicles: Small (~20 – 200 nm diameter) spherical structures enclosed by a lipid bilayer. In Gram-negative cells, they are thought to be derived from the outer membrane.

Microbial loop: The network of interactions among microorganisms at the base of the marine food web through which carbon and other nutrients move, before they are supplied to larger organisms.

Mixotroph: refers to an organism that is capable of using multiple metabolic modes for acquiring energy or carbon for growth. In the context of this Review, this refers to organisms that can use both CO_2 (autotrophy) and organic carbon (heterotrophy).

Salp: A marine tunicate that consumes plankton through filter feeding.

Ocean stratification: The division of the water column into low-density and high-density zones, with a boundary layer (the thermocline) defined by a gradient of densities across which water will not passively mix. Changes in density that lead to stratification are typically due to differences in temperature and salinity.

Oxygen minimum zone: Subsurface ocean regions that are deficient in oxygen due to poor ventilation and high rates of respiration.

Gyre: Ocean systems bounded by circular rotating winds and currents. The 5 major ocean gyres are found in the North Atlantic, North Pacific, South Atlantic, South Pacific, and Indian Oceans.

Cyanophage: A phage that infects cyanobacteria.

Paralogous genes: A pair of similar genes that were created by a duplication event.

Lysogenic phage: A bacteriophage whose genome is integrated within the host genome and replicates with it without killing the cell.

Axenic: A pure culture (ie containing only a single strain), free of any other contaminating organism.

Siderophore: A molecule that can bind iron; often used by microbes to help them obtain iron from the environment.

ToC blurb
The marine cyanobacterium *Prochlorococcus* is the most abundant photosynthetic organism on the planet. In this Review, Chisholm and colleagues highlight the enormous genomic diversity of this phototroph, discuss the factors that contribute to this diversity and consider its ecological consequences.

Author biographies
Steven J. Biller is a postdoctoral associate in the laboratory of Dr. Sallie W. Chisholm at the Massachusetts Institute of Technology, USA. He received his Ph.D. in Biological Sciences from Stanford University, USA. His current postdoctoral research investigates diverse aspects of *Prochlorococcus* ecology, physiology, and genomics.

Paul M. Berube is a research associate in the laboratory of Dr. Sallie W. Chisholm at the Massachusetts Institute of Technology, USA. He received his Ph.D. in Microbiology from the University of Washington, USA, in the laboratory of Dr. David A. Stahl, where his work focused on the starvation response of ammonia oxidizing bacteria. His current research is aimed at understanding the genomics, physiology, and ecological implications of nitrate assimilation by *Prochlorococcus*.

Debbie Lindell is an Associate Professor in the Faculty of Biology at the Technion – Israel Institute of Technology. She received her PhD from the Hebrew University of Jerusalem in 2001 where she investigated nitrogen utilization in marine cyanobacteria. She carried out post-doctoral research at MIT working on *Prochlorococcus* and their phages. She joined the Technion Faculty in 2006 and the research in her lab focuses on understanding the impact of host-phage interactions on the physiology, ecology and evolution of both cyanobacteria and cyanophages.

Sallie W. Chisholm is Lee and Geraldine Martin Professor of Environmental Studies at MIT where she holds a joint appointment in the Departments of Civil and Environmental Engineering and Biology. She received her PhD from SUNY Albany in 1974, and did post-doctoral work at Scripps Institution of Oceanography before she joined the MIT Faculty in 1976. For the past 30 years she has worked on developing *Prochlorococcus* as a model system for advancing our understanding of microbial ecology and ocean ecosystems.

Online summary

- *Prochlorococcus* is the numerically dominant phototroph in the oceans and is responsible for a notable fraction of global photosynthesis
- *Prochlorococcus* populations contain distinct subgroups with remarkable genetic and physiological diversity that contribute to its stability, abundance and broad distribution in the oceans
- Cells exhibit distinct adaptations to environmental factors such as light levels, temperature, and nutrient levels, in combinatorial arrangements
- While each individual cell has a small, “streamlined’ genome, collectively the global *Prochlorococcus* population contains a vast number of different genes
- Interactions with phages and heterotrophs have played important roles in shaping *Prochlorococcus* physiology and diversity

- *Prochlorococcus* represents a useful model system for understanding microbial ecology

References

1. Chisholm, S. *et al.* A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334,** 340–343 (1988).
2. Morel, A., Ahn, Y., Partensky, F., Vaulot, D. & Claustre, H. *Prochlorococcus* and *Synechococcus*: A comparative study of their optical properties in relation to their size and pigmentation. *J Mar Res* **51,** 617–649 (1993).
3. Flombaum, P. *et al.* Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences* **110,** 9824–9829 (2013).

   An extensive synthesis of the global distributions of the marine picocyanobacteria, including projections about how climate change may impact their abundance and habitats.

4. Schattenhofer, M. *et al.* Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environmental Microbiology* **11,** 2078–2093 (2009).
5. Partensky, F., Blanchot, J. & Vaulot, D. Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bulletin de l'Institut oceanographique, Monaco* **19,** 457–476 (1999).
6. Dufresne, A. *et al.* Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proceedings of the National Academy of Sciences* **100,** 10020–10025 (2003).
7. Rocap, G. *et al.* Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424,** 1042–1047 (2003).

   A detailed comparison of genomes from a representative HL- and LL-adapted strain reveals many of the fundamental genomic distinctions that correlate with their different physiologies and evolutionary histories. This work also highlights the power of comparative genomics in microbial ecology.

8. Goericke, R. & Repeta, D. J. The pigments of *Prochlorococcus marinus*: the presence of divinyl chlorophyll *a* and *b* in a marine procaryote. *Limnology and Oceanography* **37,** 425–433 (1992).
9. Moore, L., Goericke, R. & Chisholm, S. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Marine Ecology Progress Series* **116,** 259–275 (1995).
10. Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63,** 106–127 (1999).
11. Partensky, F. & Garczarek, L. *Prochlorococcus*: Advantages and limits of minimalism. *Annu. Rev. Marine. Sci.* **2,** 305–331 (2010).
12. Huston, M. A. & Wolverton, S. The global distribution of net primary production: resolving the paradox. *Ecological Monographs* **79,** 343–377 (2009).
13. Olson, R. J., Chisholm, S., Zettler, E. R., Altabet, M. & Dusenberry, J. A. Spatial and temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. *Deep Sea Research Part A, Oceanographic Research Papers* **37,** 1033–1051 (1990).
14. Vaulot, D., Marie, D., Olson, R. J. & Chisholm, S. W. Growth of *Prochlorococcus*, a

Photosynthetic Prokaryote, in the Equatorial Pacific Ocean. *Science* **268,** 1480–1482 (1995).
15. Holtzendorff, J. *et al.* Diel expression of cell cycle-related genes in synchronized cultures of *Prochlorococcus* sp. strain PCC 9511. *Journal of Bacteriology* **183,** 915–920 (2001).
16. Holtzendorff, J. *et al.* Synchronized expression of *ftsZ* in natural *Prochlorococcus* populations of the Red Sea. *Environmental Microbiology* **4,** 644–653 (2002).
17. Zinser, E. R. *et al.* Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus*. *PLoS ONE* **4,** e5135 (2009).
18. Waldbauer, J. R., Rodrigue, S., Coleman, M. L. & Chisholm, S. W. Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PLoS ONE* **7,** e43432 (2012).
19. Ottesen, E. A. *et al.* Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* **345,** 207–212 (2014).
20. Simon, M., Grossart, H.-P., Schweitzer, B. & Ploug, H. Microbial ecology of organic aggregates in aquatic ecosystems. *Aquatic Microbial Ecology* **28,** 175–211 (2002).
21. Malfatti, F. & Azam, F. Atomic force microscopy reveals microscale networks and possible symbioses among pelagic marine bacteria. *Aquatic Microbial Ecology* **58,** 1–14 (2009).
22. Bertilsson, S., Berglund, O., Karl, D. & Chisholm, S. Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea. *Limnology and Oceanography* **48,** 1721–1731 (2003).
23. Heldal, M., Scanlan, D. J., Norland, S., Thingstad, F. & Mann, N. H. Elemental composition of single cells of various strains of marine *Prochlorococcus* and *Synechococcus* using X-ray microanalysis. *Limnology and Oceanography* **48,** 1732–1743 (2003).
24. Grob, C. *et al.* Elemental composition of natural populations of key microbial groups in Atlantic waters. *Environmental Microbiology* **15,** 3054–3064 (2013).
25. Van Mooy, B. A. S., Rocap, G., Fredricks, H. F., Evans, C. T. & Devol, A. H. Sulfolipids dramatically decrease phosphorus demand by picocyanobacteria in oligotrophic marine environments. *Proceedings of the National Academy of Sciences* **103,** 8607–8612 (2006).

This study highlights the strong selective pressure that phosphorus-limitation has imposed on the evolution of *Prochlorococcus.*

26. Zwirglmaier, K. *et al.* Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environmental Microbiology* **10,** 147–161 (2008).
27. Moore, L. R., Rocap, G. & Chisholm, S. W. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393,** 464–467 (1998).

This study demonstrates that genetically distinct *Prochlorococcus* strains isolated from the same water sample have distinct light adaptations. This set the stage for the development of *Prochlorococcus* as a model system that could be used to link field observations with physiological properties that are determined through the study of laboratory cultures.

28. Scanlan, D. J. & West, N. J. Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiol Ecol* **40,** 1–12 (2002).
29. Ahlgren, N. A., Rocap, G. & Chisholm, S. W. Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environmental Microbiology* **8,** 441–454 (2006).
30. Zinser, E. R. *et al.* Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnology and Oceanography* **52,** 2205–2220 (2007).
31. West, N. J. & Scanlan, D. J. Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the Eastern North Atlantic Ocean. *Applied and Environmental Microbiology* **65,** 2585–2591 (1999).

This article provides the first description of how different *Prochlorococcus* light ecotypes partition in the water column.

32. West, N. J. *et al.* 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 (2001).
33. Johnson, Z. I. *et al.* Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311,** 1737–1740 (2006).

This study revealed temperature as a factor correlated with the distribution of high-light adapted *Prochlorococcus* ecotypes along ocean gradients, and provided evidence that the physiology of cells in culture matches their distributions in the wild.

34. Zwirglmaier, K. *et al.* Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environmental Microbiology* **9,** 1278–1290 (2007).
35. Malmstrom, R. R. *et al.* Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *The ISME Journal* **4,** 1252–1264 (2010).

Using data from two long-term ocean time-series stations, this paper highlights the remarkable reproducibility of *Prochlorococcus* ecotype abundances over many years.

36. Martiny, A. C., Tai, A. P. K., Veneziano, D., Primeau, F. & Chisholm, S. W. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environmental Microbiology* **11,** 823–832 (2009).
37. Rocap, G., Distel, D. L., Waterbury, J. B. & Chisholm, S. W. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Applied and Environmental Microbiology* **68,** 1180–1191 (2002).
38. Ferris, M. J. & Palenik, B. Niche adaptation in ocean cyanobacteria. *Nature* **396,** 226–228 (1998).
39. Jameson, E., Joint, I., Mann, N. H. & Mühling, M. Application of a novel *rpoC1*-RFLP approach reveals that marine *Prochlorococcus* populations in the Atlantic gyres are

composed of greater microdiversity than previously described. *Microbial Ecology* **55,** 141–151 (2008).
40. Urbach, E., Scanlan, D. J., Distel, D. L., Waterbury, J. B. & Chisholm, S. W. Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria). *J Mol Evol* **46,** 188–201 (1998).
41. Penno, S., Lindell, D. & Post, A. F. Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environmental Microbiology* **8,** 1200–1211 (2006).
42. Mühling, M. M. On the culture-independent assessment of the diversity and distribution of *Prochlorococcus*. *Environmental Microbiology* **14,** 567–579 (2012).
43. Urbach, E., Robertson, D. L. & Chisholm, S. W. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. *Nature* **355,** 267–270 (1992).
44. Kettler, G. C. *et al.* Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genetics* **3,** e231 (2007).
45. Kashtan, N. *et al.* Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* **344,** 416–420 (2014).

This study demonstrates the vast genomic diversity of *Prochorococcus* cells in a single sample of seawater, and argues that hundreds of diverse subpopulations contribute to the dynamics and stability of the global *Prochlorococcus* federation.

46. Partensky, F., Hoepffner, N., Li, W. & Ulloa, O. Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) strains isolated from the North Atlantic and the Mediterranean Sea. *Plant Physiology* **101,** 285–296 (1993).
47. Huang, S. *et al.* Novel lineages of *Prochlorococcus* and *Synechococcus* in the global oceans. *The ISME Journal* **6,** 285–297 (2012).
48. Malmstrom, R. R. *et al.* Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *The ISME Journal* **7,** 184–198 (2013).
49. Rusch, D. B., Martiny, A. C., Dupont, C. L., Halpern, A. L. & Venter, J. C. Characterization of *Prochlorococcus* clades from iron-depleted oceanic regions. *Proceedings of the National Academy of Sciences* **107,** 16184–16189 (2010).

This study demonstrates the utility of metagenomic data for characterizing the distribution and key features of previously unknown and uncultured lineages of *Prochlorococcus*.

50. West, N. J., Lebaron, P., Strutton, P. G. & Suzuki, M. T. A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *The ISME Journal* **5,** 933–944 (2011).
51. Shi, Y., Tyson, G. W., Eppley, J. M. & Delong, E. F. Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME Journal* **5,** 999–1013 (2011).
52. Zinser, E. R. *et al. Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Applied and Environmental Microbiology* **72,** 723–732 (2006).

53. Coleman, M. L. & Chisholm, S. W. Code and context: *Prochlorococcus* as a model for cross-scale biology. *Trends Microbiol* **15,** 398–407 (2007).
54. He, Q., Dolganov, N., Bjorkman, O. & Grossman, A. R. The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. *The Journal Of Biological Chemistry* **276,** 306–314 (2001).
55. Li, B. *et al.* Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proceedings of the National Academy of Sciences* **107,** 10430–10435 (2010).
56. Lavin, P., González, B., Santibáñez, J. F., Scanlan, D. J. & Ulloa, O. Novel lineages of *Prochlorococcus* thrive within the oxygen minimum zone of the eastern tropical South Pacific. *Environmental Microbiology Reports* **2,** 728–738 (2010).
57. Giovannoni, S. J., Thrash, J. C. & Ben Temperton. Implications of streamlining theory for microbial ecology. *The ISME Journal* **8,** 1553–1565 (2014).
58. Scanlan, D. J. *et al.* Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* **73,** 249–299 (2009).

This extensive review examines the similarities and differences among *Synechococcus* and *Prochlorococcus* genomes from an environmental perspective.

59. Sun, Z. & Blanchard, J. L. Strong Genome-Wide Selection Early in the Evolution of *Prochlorococcus* Resulted in a Reduced Genome through the Loss of a Large Number of Small Effect Genes. *PLoS ONE* **9,** e88837 (2014).
60. Biller, S. J. *et al.* Genomes of diverse isolates of the marine cyanobacterium *Prochlorococcus*. *Scientific Data* **1,** 140034 (2014).
61. Baumdicker, F., Hess, W. R. & Pfaffelhuber, P. The infinitely many genes model for the distributed genome of bacteria. *Genome Biol Evol* **4,** 443–456 (2012).
62. Coleman, M. L. *et al.* Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311,** 1768–1770 (2006).

This study reveals the importance of genomic islands as a hotspot for the integration of ecologically important flexible genes into *Prochlorococcus* genomes.

63. Humbert, J.-F. *et al.* A Tribute to Disorder in the Genome of the Bloom-Forming Freshwater Cyanobacterium *Microcystis aeruginosa*. *PLoS ONE* **8,** e70747 (2013).
64. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304,** 66–74 (2004).
65. Palenik, B. *et al.* The genome of a motile marine *Synechococcus*. *Nature* **424,** 1037–1042 (2003).
66. Dufresne, A. *et al.* Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9,** R90 (2008).
67. Luo, H., Friedman, R., Tang, J. & Hughes, A. L. Genome reduction by deletion of paralogs in the marine cyanobacterium *Prochlorococcus*. *Mol Biol Evol* **28,** 2751–2760 (2011).
68. Martiny, A. C., Coleman, M. L. & Chisholm, S. W. Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proceedings of the National Academy of Sciences* **103,** 12552–12557 (2006).

This article shows that P-limitation is one of the strongest selective pressures shaping gene content of *Prochlorococcus* in the Atlantic versus the Pacific Ocean.

69. Coleman, M. L. & Chisholm, S. W. Ecosystem-specific selection pressures revealed through comparative population genomics. *Proceedings of the National Academy of Sciences* **107,** 18634–18639 (2010).
70. Thompson, L. R. *et al.* Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences* **108,** E757–E764 (2011).
71. Kelly, L., Ding, H., Huang, K. H., Osburne, M. S. & Chisholm, S. W. Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *The ISME Journal* **7,** 1827–1841 (2013).
72. Avrani, S., Wurtzel, O., Sharon, I., Sorek, R. & Lindell, D. Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* **474,** 604–608 (2011).

This study highlights the role of genetic diversity in genomic islands in maintaining the coexistence of *Prochlorococcus* and cyanophages.

73. Collins, R. E. & Higgs, P. G. Testing the Infinitely Many Genes model for the evolution of the bacterial core genome and pangenome. *Mol Biol Evol* **29,** 3413–3425 (2012).
74. Kislyuk, A. O., Haegeman, B., Bergman, N. H. & Weitz, J. S. Genomic fluidity: an integrative view of gene diversity within microbial populations. *BMC Genomics* **12,** 32 (2011).
75. Moore, L. & Chisholm, S. Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnology and Oceanography* **44,** 628–638 (1999).
76. Dufresne, A., Garczarek, L. & Partensky, F. Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol* **6,** R14 (2005).
77. Osburne, M. S., Holmbeck, B. M., Coe, A. & Chisholm, S. W. The spontaneous mutation frequencies of Prochlorococcus strains are commensurate with those of other bacteria. *Environmental Microbiology Reports* **3,** 744–749 (2011).
78. Hu, J. & Blanchard, J. L. Environmental sequence data from the Sargasso Sea reveal that the characteristics of genome reduction in *Prochlorococcus* are not a harbinger for an escalation in genetic drift. *Mol Biol Evol* **26,** 5–13 (2008).
79. Cohan, F. M. Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361,** 1985–1996 (2006).
80. Rodriguez-Valera, F. *et al.* Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7,** 828–836 (2009).
81. Cordero, O. X. & Polz, M. F. Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol* **12,** 263–273 (2014).
82. Rodriguez-Valera, F. & Ussery, D. W. Is the pan-genome also a pan-selectome? *F1000Res* **1,** 16 (2012).
83. Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424,** 1047–1051 (2003).

84. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol* **3,** e144 (2005).
85. Sullivan, M. B. *et al.* The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environmental Microbiology* **11,** 2935–2951 (2009).
86. Sullivan, M. B. *et al.* Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environmental Microbiology* **12,** 3035–3056 (2010).
87. Labrie, S. J. *et al.* Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environmental Microbiology* **15,** 1356–1376 (2013).
88. Parsons, R. J., Breitbart, M., Lomas, M. W. & Carlson, C. A. Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. *The ISME Journal* **6,** 273–284 (2012).
89. Williams, K. P. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res* **30,** 866–875 (2002).
90. Lindell, D. *et al.* Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449,** 83–86 (2007).
91. Lindell, D. *et al.* Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proceedings of the National Academy of Sciences* **101,** 11013–11018 (2004).
92. Zeidner, G. *et al.* Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environmental Microbiology* **7,** 1505–1513 (2005).
93. Sullivan, M. B. *et al.* Prevalence and evolution of core photosystem II genes inmMarine cyanobacterial viruses and their hosts. *PLoS Biol* **4,** e234 (2006).
94. Cai, F., Axen, S. D. & Kerfeld, C. A. Evidence for the widespread distribution of CRISPR-Cas system in the Phylum *Cyanobacteria*. *RNA Biology* **10,** 1–7 (2013).
95. Weinberger, A. D., Wolf, Y. I., Lobkovsky, A. E., Gilmore, M. S. & Koonin, E. V. Viral diversity threshold for adaptive immunity in prokaryotes. *mBio* **3,** e00456–e00412 (2012).
96. Avrani, S., Schwartz, D. & Lindell, D. Virus-host swinging party in the oceans: Incorporating biological complexity into paradigms of antagonistic coexistence. *Mob Genet Elements* **2,** 88–95 (2012).
97. Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Bacterial photosynthesis genes in a virus. *Nature* **424,** 741 (2003).

This paper is the first to report the presence of photosynthesis genes in a virus.

98. Millard, A. D., Zwirglmaier, K., Downey, M. J., Mann, N. H. & Scanlan, D. J. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environmental Microbiology* **11,** 2370–2387 (2009).
99. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438,** 86–89 (2005).

100. Zeng, Q. & Chisholm, S. W. Marine Viruses Exploit Their Host's Two-Component Regulatory System in Response to Resource Limitation. *Curr Biol* **22,** 124–128 (2012).
101. Carini, P., Steindler, L., Beszteri, S. & Giovannoni, S. J. Nutrient requirements for growth of the extreme oligotroph '*Candidatus* Pelagibacter ubique' HTCC1062 on a defined medium. *The ISME Journal* **7,** 592–602 (2013).
102. Bertilsson, S., Berglund, O., Pullin, M. & Chisholm, S. Release of dissolved organic matter by *Prochlorococcus*. *Vie et Milieu* **55,** 225–232 (2005).
103. Chisholm, S. W. *et al. Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b*. *Arch Microbiol* **157,** 297–300 (1992).
104. Rippka, R. *et al. Prochlorococcus marinus* Chisholm *et al.* 1992 subsp. *pastoris* subsp. nov. strain PCC 9511, the first axenic chlorophyll a2/b2-containing cyanobacterium (*Oxyphotobacteria*). *International Journal of Systematic and Evolutionary Microbiology* **50,** 1833–1847 (2000).
105. Saito, M., Moffett, J., Chisholm, S. & Waterbury, J. Cobalt limitation and uptake in *Prochlorococcus*. *Limnology and Oceanography* **47,** 1629–1636 (2002).
106. Berube, P. M. *et al.* Physiology and evolution of nitrate acquisition in *Prochlorococcus*. *The ISME Journal* ***in press*** (2014).
107. Morris, J. J., Kirkegaard, R., Szul, M. J., Johnson, Z. I. & Zinser, E. R. Facilitation of robust growth of *Prochlorococcus* colonies and dilute liquid cultures by 'helper' heterotrophic bacteria. *Applied and Environmental Microbiology* **74,** 4530–4534 (2008).
108. Sher, D., Thompson, J. W., Kashtan, N., Croal, L. & Chisholm, S. W. Response of *Prochlorococcus* ecotypes to co-culture with diverse marine bacteria. *The ISME Journal* **5,** 1125–1132 (2011).
109. Morris, J. J., Johnson, Z. I., Szul, M. J., Keller, M. & Zinser, E. R. Dependence of the cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at the ocean's surface. *PLoS ONE* **6,** e16805 (2011).

Experimental demonstration of the importance of heterotroph interactions for *Prochlorococcus* growth in the wild.

110. Morris, J. J., Lenski, R. E. & Zinser, E. R. The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. *mBio* **3,** e00036–12 (2012).
111. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30,** 108–160 (2013).
112. Biller, S. J. *et al.* Bacterial Vesicles in Marine Ecosystems. *Science* **343,** 183–186 (2014).
113. Tettelin, H., Riley, D., Cattuto, C. & Medini, D. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol* **11,** 472–477 (2008).
114. Doolittle, W. F. & Zhaxybayeva, O. Metagenomics and the Units of Biological Organization. *BioScience* **60,** 102–112 (2010).
115. Becker, J. W. *et al.* Closely related phytoplankton species produce similar suites of dissolved organic matter. *Front Microbiol* **5,** 1–14 (2014).
116. Azam, F. & Malfatti, F. Microbial structuring of marine ecosystems. *Nat Rev Microbiol* **5,** 782–791 (2007).

117. Goericke, R., Strom, S. L. & Bell, R. A. Distribution and sources of cyclic pheophorbides in the marine environment. *Limnology and Oceanography* **45,** 200–211 (2000).
118. Sutherland, K. R., Madin, L.P. & Stocker, R. Filtration of submicrometer particles by pelagic tunicates. *Proceedings of the National Academy of Sciences* **107,** 15129–15134 (2010).
119. Christaki, U., Jacquet, S., Dolan, J. R., Vaulot, D. & Rassoulzadegan, F. Growth and grazing on *Prochlorococcus* and *Synechococcus* by two marine ciliates. *Limnology and Oceanography* **44,** 52–61 (1999).
120. Hirose, M., Katano, T. & Nakano, S.-I. Growth and grazing mortality rates of *Prochlorococcus*, *Synechococcus* and eukaryotic picophytoplankton in a bay of the Uwa Sea, Japan. *Journal of Plankton Research* **30,** 241–250 (2008).
121. Guillou, L., Jacquet, S., Chretiennot-Dinet, M.-J. & Vaulot, D. Grazing impact of two small heterotrophic flagellates on *Prochlorococcus* and *Synechococcus*. *Aquatic Microbial Ecology* **26,** 201–207 (2001).
122. Hartmann, M., Zubkov, M. V., Scanlan, D. J. & Lepère, C. *In situ* interactions between photosynthetic picoeukaryotes and bacterioplankton in the Atlantic Ocean: evidence for mixotrophy. *Environmental Microbiology Reports* **5,** 835–840 (2013).
123. Frias-Lopez, J., Thompson, A., Waldbauer, J. & Chisholm, S. W. Use of stable isotope-labelled cells to identify active grazers of picocyanobacteria in ocean surface waters. *Environmental Microbiology* **11,** 512–525 (2009).
124. Raven, J. A., Beardall, J., Flynn, K. J. & Maberly, S. C. Phagotrophy in the origins of photosynthesis in eukaryotes and as a complementary mode of nutrition in phototrophs: relation to Darwin's insectivorous plants. *J Exp Bot* **60,** 3975–3987 (2009).
125. Richardson, T. L. & Jackson, G. A. Small phytoplankton and carbon export from the surface ocean. *Science* **315,** 838–840 (2007).
126. Zubkov, M. V., Fuchs, B. M., Tarran, G. A., Burkill, P. H. & Amann, R. High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Applied and Environmental Microbiology* **69,** 1299–1304 (2003).
127. Gómez-Pereira, P. R. *et al.* Comparable light stimulation of organic nutrient uptake by SAR11 and *Prochlorococcus* in the North Atlantic subtropical gyre. *The ISME Journal* **7,** 603–614 (2013).
128. Mary, I. *et al.* Light enhanced amino acid uptake by dominant bacterioplankton groups in surface waters of the Atlantic Ocean. *FEMS Microbiol Ecol* **63,** 36–45 (2008).
129. Michelou, V. K., Cottrell, M. T. & Kirchman, D. L. Light-stimulated bacterial production and amino acid assimilation by cyanobacteria and other microbes in the North Atlantic ocean. *Applied and Environmental Microbiology* **73,** 5539–5546 (2007).
130. Del Carmen Muñoz-Marín, M. *et al. Prochlorococcus* can use the Pro1404 transporter to take up glucose at nanomolar concentrations in the Atlantic Ocean. *Proceedings of the National Academy of Sciences* **110,** 8597–8602 (2013).

This article demonstrates the potential for *Prochlorococcus* photoheterotrophic growth in the wild.

131. Gómez-Baena, G. *et al.* Glucose uptake and its effect on gene expression in

Prochlorococcus. *PLoS ONE* **3,** e3416 (2008).
132. Zhaxybayeva, O., Doolittle, W. F., Papke, R. T. & Gogarten, J. P. Intertwined evolutionary histories of marine *Synechococcus* and *Prochlorococcus marinus*. *Genome Biol Evol* **1,** 325–339 (2009).
133. Ting, C. S., Rocap, G., King, J. & Chisholm, S. W. Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol* **10,** 134–142 (2002).
134. Mackey, K. R. M. *et al.* Effect of Temperature on Photosynthesis and Growth in Marine *Synechococcus* spp. *Plant Physiology* **163,** 815–829 (2013).
135. Pittera, J. *et al.* Connecting thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. *The ISME Journal* **8,** 1221–1236 (2014).
136. Mann, E., Ahlgren, N., Moffett, J. & Chisholm, S. Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnology and Oceanography* **47,** 976–988 (2002).
137. Chen, B., Liu, H., Landry, M. R., Chen, M. & Sun, J. Estuarine nutrient loading affects phytoplankton growth and microzooplankton grazing at two contrasting sites in Hong Kong coastal waters. *Marine Ecology Progress Series* **379,** 77–90 (2009).
138. Moore, L. *et al.* Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol. Oceanogr.: Methods* **5,** 353–362 (2007).
139. Martiny, A. C., Huang, Y. & Li, W. Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environmental Microbiology* **11,** 1340–1347 (2009).
140. Feingersch, R. *et al.* Potential for phosphite and phosphonate utilization by *Prochlorococcus*. *The ISME Journal* **6,** 827–834 (2012).
141. Martinez, A., Tyson, G. W. & Delong, E. F. Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. *Environmental Microbiology* **12,** 222–238 (2010).
142. Martinez, A., Osburne, M. S., Sharma, A. K., Delong, E. F. & Chisholm, S. W. Phosphite utilization by the marine picocyanobacterium *Prochlorococcus* MIT9301. *Environmental Microbiology* **14,** 1363–1377 (2012).
143. Grzymski, J. J. & Dussaq, A. M. The significance of nitrogen cost minimization in proteomes of marine microorganisms. *The ISME Journal* **6,** 71–80 (2012).
144. Bragg, J. G. & Hyder, C. L. Nitrogen versus carbon use in prokaryotic genomes and proteomes. *Proc Biol Sci* **271 Suppl 5,** S374–7 (2004).
145. Gilbert, J. D. & Fagan, W. F. Contrasting mechanisms of proteomic nitrogen thrift in *Prochlorococcus*. *Molecular Ecology* **20,** 92–104 (2011).
146. Garcia-Fernandez, J. M., de Marsac, N. T. & Diez, J. Streamlined regulation and gene loss as adaptive mechanisms in *Prochlorococcus* for optimized nitrogen utilization in oligotrophic environments. *Microbiology and Molecular Biology Reviews* **68,** 630–638 (2004).
147. Martiny, A. C., Kathuria, S. & Berube, P. M. Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proceedings of the National Academy of Sciences* **106,** 10787–10792 (2009).
148. Kamennaya, N. A. & Post, A. F. Characterization of cyanate metabolism in marine *Synechococcus* and *Prochlorococcus* spp. *Applied and Environmental Microbiology* **77,** 291–301 (2011).
149. Moore, L., Post, A., Rocap, G. & Chisholm, S. Utilization of different nitrogen sources

by the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Limnology and Oceanography* **47,** 989–996 (2002).
150. Thompson, A. W., Huang, K., Saito, M. A. & Chisholm, S. W. Transcriptome response of high- and low-light-adapted *Prochlorococcus* strains to changing iron availability. *The ISME Journal* **5,** 1580–1594 (2011).