



Unique Patterns and Biogeochemical Relevance of Two-Component Sensing in Marine Bacteria

Noelle A. Held,^a Matthew R. McIlvin,^b Dawn M. Moran,^b Michael T. Laub,^{c,d} Mak A. Saito^b

^aMIT-WHOI Joint Program in Oceanography/Applied Ocean Science and Engineering, Woods Hole, Massachusetts, USA

^bMarine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA

^cDepartment of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

^dHoward Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

ABSTRACT Two-component sensory (TCS) systems link microbial physiology to the environment and thus may play key roles in biogeochemical cycles. In this study, we surveyed the TCS systems of 328 diverse marine bacterial species. We identified lifestyle traits such as copiotrophy and diazotrophy that are associated with larger numbers of TCS system genes within the genome. We compared marine bacterial species with 1,152 reference bacterial species from a variety of habitats and found evidence of extra response regulators in marine genomes. Examining the location of TCS genes along the circular bacterial genome, we also found that marine bacteria have a large number of “orphan” genes, as well as many hybrid histidine kinases. The prevalence of “extra” response regulators, orphan genes, and hybrid TCS systems suggests that marine bacteria break with traditional understanding of how TCS systems operate. These trends suggest prevalent regulatory networking, which may allow coordinated physiological responses to multiple environmental signals and may represent a specific adaptation to the marine environment. We examine phylogenetic and lifestyle traits that influence the number and structure of two-component systems in the genome, finding, for example, that a lack of two-component systems is a hallmark of oligotrophy. Finally, in an effort to demonstrate the importance of TCS systems to marine biogeochemistry, we examined the distribution of *Prochlorococcus/Synechococcus* response regulator PMT9312_0717 in metaproteomes of the tropical South Pacific. We found that this protein’s abundance is related to phosphate concentrations, consistent with a putative role in phosphate regulation.

IMPORTANCE Marine microbes must manage variation in their chemical, physical, and biological surroundings. Because they directly link bacterial physiology to environmental changes, TCS systems are crucial to the bacterial cell. This study surveyed TCS systems in a large number of marine bacteria and identified key phylogenetic and lifestyle patterns in environmental sensing. We found evidence that, in comparison with bacteria as a whole, marine organisms have irregular TCS system constructs which might represent an adaptation specific to the marine environment. Additionally, we demonstrate the biogeochemical relevance of TCS systems by correlating the presence of the PMT9312_0717 response regulator protein to phosphate concentrations in the South Pacific. We highlight that despite their potential ecological and biogeochemical relevance, TCS systems have been understudied in the marine ecosystem. This report expands our understanding of the breadth of bacterial TCS systems and how marine bacteria have adapted to survive in their unique environment.

KEYWORDS biogeochemistry, cell signaling, gene regulation, marine microbiology, proteomics, regulatory network, two-component system

Citation Held NA, McIlvin MR, Moran DM, Laub MT, Saito MA. 2019. Unique patterns and biogeochemical relevance of two-component sensing in marine bacteria. *mSystems* 4:e00317-18. <https://doi.org/10.1128/mSystems.00317-18>.

Editor Olivia Mason, Florida State University

Copyright © 2019 Held et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Mak A. Saito, msaito@whoi.edu.

Received 11 December 2018

Accepted 11 January 2019

Published 5 February 2019

A bacterium's survival is dependent on its ability to respond to changes in its environment. This is especially true for marine microbes, which experience changes in nutrient availability, light, temperature, and community structure that can occur on time scales as short as hours (1). Two-component sensory (TCS) systems, which modulate gene expression on short time scales, are the most common sensory systems in prokaryotes (2). The organism's complement of two-component-system genes may thus reveal details about its lifestyle, ecological niche, and physiological complexity.

Canonically, two-component sensory systems are composed of a single histidine kinase and response regulator protein pair (Fig. 1). The histidine kinase contains a sensory domain that is activated by a specific stimulus, which may be a small molecule, nutrient, or physical property such as light or temperature. Upon activation, the histidine kinase is autophosphorylated on a conserved histidine residue, inducing conformational changes that enhance interactions with the response regulator protein (3). Through a highly specific protein-protein interaction, the phosphate moiety is transferred from the histidine kinase to a conserved aspartate residue on the response regulator (4). This typically stimulates binding of the response regulator to a DNA promoter region, resulting in transcription of genes in the downstream operon (5). A variation on this model is the hybrid histidine kinase, in which the kinase and response regulator are located on a single protein. An intermediary protein or set of proteins is involved in transmitting the signal, allowing multiple levels of control and a fine-tuned physiological response (3).

Two-component sensory systems represent a direct link between the environment and the physiology of the prokaryotic cell and, as such, may be important players in biogeochemical cycles. One well-studied example is the Pho system, composed of the histidine kinase PhoR and response regulator PhoB. The Pho system is common in marine bacteria and regulates genes that are involved in phosphate acquisition (6). In *Prochlorococcus*, activation of PhoR by low intracellular phosphate stimulates transcription of alkaline phosphatase. Alkaline phosphatase cleaves phosphate from organic matter, providing a source of phosphate that would otherwise be inaccessible (7). As the vector linking microbial physiology (i.e., the expression of alkaline phosphatase enzyme) to ocean chemistry (i.e., phosphate availability), the Pho system may be a key regulator of phosphorus cycling in the ocean. Other two-component sensory (TCS) systems may also play important roles in mediating microbe-environment interactions, though we do not yet understand their biogeochemical contexts.

In this study, we surveyed the TCS system genes of 328 diverse marine bacteria, identifying phylogenetic and lifestyle factors that correlate with greater numbers of sensory genes. We compared these marine bacteria to a curated reference collection of 1,152 bacterial genomes, most derived from the GEBA initiative (8). This allowed us to identify key differences in how TCS systems are structured in marine bacteria versus bacteria in general. To demonstrate the importance of TCS systems to marine biogeochemistry, we examined the distribution of a putative phosphate-sensing *Prochlorococcus* response regulator (PMT9312_0717) in metaproteomes from the tropical Pacific. We highlight gaps in our knowledge of marine TCS and emphasize the importance of TCS to our overall understanding of marine bacteria.

RESULTS

Lifestyle influences on TCS gene abundance. We examined the genomes of 328 diverse marine bacteria publicly available in the JGI IMG data warehouse (see Table S1 in the supplemental material). The data set emphasizes cultivable and oceanographically important organisms such as *Prochlorococcus*, *Synechococcus*, *Pelagibacter*, *Alteromonas*, and *Roseobacter*. In total, 15 phyla and 183 genera are represented from a variety of habitats, including coastal ecosystems, the open ocean, hydrothermal vent systems, host-associated environments, and marine sediments. All genomes were labeled high-quality finished or permanent drafts. TCS system genes were identified

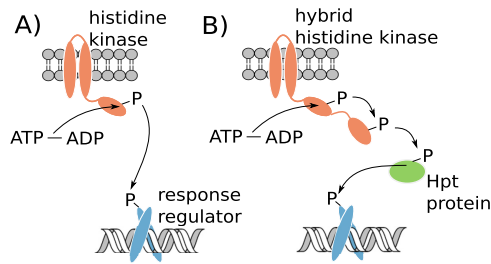


FIG 1 Overview of two-component-system signaling in (A) a traditional histidine kinase–response regulator system and (B) a hybrid histidine kinase system. In panel A, the phosphorylation is transferred from the histidine kinase to the response regulator by a direct protein–protein interaction. In panel B, the phosphorylation is transferred to an internal receiver domain on the histidine kinase, then to one or more histidine phosphotransfer (Hpt) proteins, and finally to the terminal response regulator.

using highly conserved protein family domains for histidine kinases or response regulators as described in Materials and Methods.

We began by examining one half of the two-component system—the sensory histidine protein kinase (HPK). The number of histidine kinases in each genome ranges from 1 (*Pelagibacter*) to 174 (*Desulfovibrio inopinatus* DSM 10711) and is related to genome size; we found on average 0.902 histidine kinases (HPK) per 100 protein-encoding genes (Fig. 2). A K-means clustering analysis was performed on the proportion of the genome devoted to histidine kinases (expressed for human readability as the number of histidine kinases per 100 protein-encoding genes) versus genome size. This analysis revealed that the number of histidine kinases per 100 protein-encoding genes is dependent on both phylogeny and lifestyle. The first cluster contains oligotrophic picoplankton such as *Prochlorococcus*, *Pelagibacter*, and open ocean *Synechococcus*. These organisms have very small genomes and few histidine kinases per 100

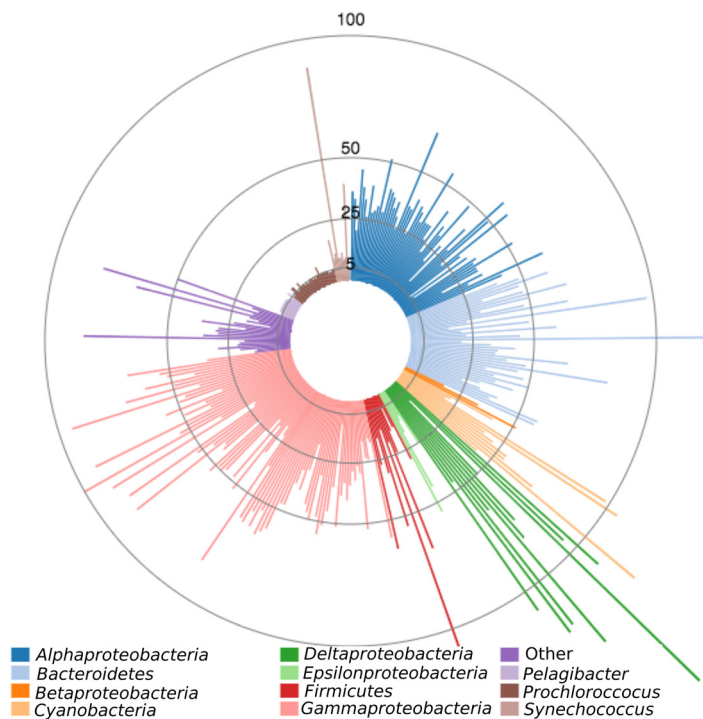


FIG 2 Number of histidine kinase sensory genes in the genomes of 328 diverse marine bacterial species (scale indicated by concentric circles). Phylogenetic groups of interest are delineated by color. The number of histidine kinases in the data set ranges from 1 (*Pelagibacter*) to 174 (*Desulfovibrio inopinatus* DSM 10711).

TABLE 1 Characteristics of copiotrophs versus oligotrophs and their TCS system genes^a

Genus	Example genome size (bp)	Example % growth rate (per day)	Lifestyle(s)	HPK/100 genes	RR/HPK ratio	% hybrid HPKs	Reference
<i>Pelagibacter</i>	1,200–1,400	0.4–0.58	Oligotroph	0.387	0.83	Typically 0	59
<i>Prochlorococcus</i>	1,200–2,000	0.51–0.83	Oligotroph	0.76	1.22	Typically 0	60
<i>Synechococcus</i>	1,500–3,000	1	Oligotroph	1.005	1.23	0–40	60
<i>Trichodesmium</i>	~5,000	0.29	Oligotroph	0.694	0.753	15–35	61
<i>Crocospaera</i>	~6,000	0.5	Oligotroph	0.723	0.992	~35	62
<i>Roseobacter</i>	~5,000	1.45	Varies/copiotroph	0.755	0.991	10–40	63
<i>Vibrio</i>	~5,000	Up to 14.3	Copiotroph	1.25	1.07	25–50	64
<i>Alteromonas</i>	4,000–4,500	6	Copiotroph	1.43	1.06	~40	65
<i>Pseudoalteromonas</i>	3,000–5,000	~30	Copiotroph	1.5	1.1	40–50	66

^aHPK, histidine kinase; RR, response regulator.

protein-encoding genes. Cluster 2 contains *Rhodobacter*, *Vibrio*, coastal *Synechococcus*, and *Alteromonas* and *Pseudoalteromonas* genomes, which are larger and have more histidine kinases per 100 protein-encoding genes. Cluster 3 is composed mainly of *Alteromonas* and *Pseudoalteromonas* genomes that have genome sizes similar to those of cluster 2 organisms but more histidine kinases per 100 protein-encoding genes. Cluster 4 organisms have the greatest numbers of histidine kinases per 100 protein-encoding genes; many have specific lifestyle traits such as particle association, parasitism, and mat formation. A number are sulfate-reducing *Deltaproteobacteria*.

Marine organisms are often classified by their nutritional preferences as copiotrophs (adapted to high-nutrient conditions), oligotrophs (organisms adapted to low-nutrient conditions), or as organisms that fall between those classifications (9–12). This provides a framework for understanding properties such as growth rate, cell size, and genome size (Table 1). We classified marine bacteria as copiotrophs or oligotrophs based on published isolation and laboratory growth conditions and found that the copiotrophs have significantly more histidine kinases per gene than the oligotrophs ($P = 3e^{-15}$ by Student's t test) (Fig. 3B).

Unusual patterns in marine TCS sensing genes. We compared the TCS system genes of marine bacteria with those of 1,152 reference bacteria (Table S2). The reference data set includes bacteria from diverse lineages and habitats, including terrestrial, freshwater, host-associated, and marine bacteria, and is representative of trends in the two-component systems of bacteria as a whole. The phylogenetic distribution of genomes in the reference data set is broad (see Fig. S1 in the supple-

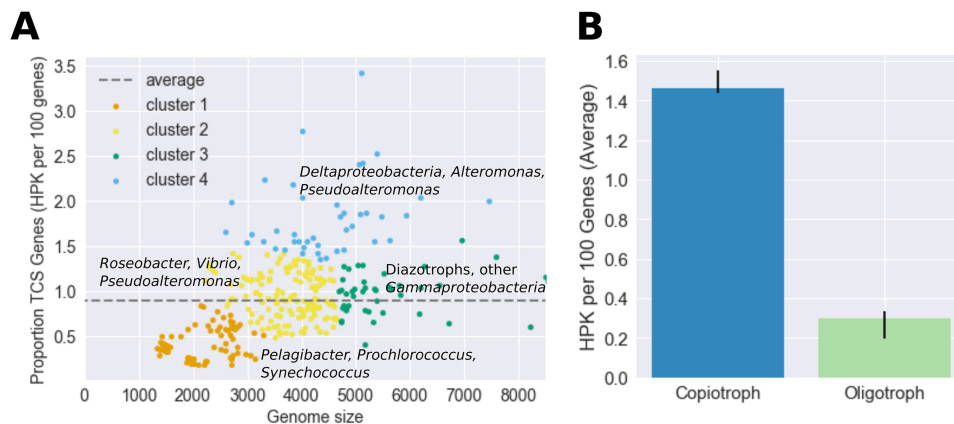


FIG 3 (A) K-means clustering of the number of histidine kinases (HPK) per 100 protein-encoding genes as a function of genome size. The dashed line represents the average, 0.902 histidine kinases per 100 protein-encoding genes in the genome. (B) The number of histidine kinases per 100 protein-encoding genes in the genome for organisms unambiguously designated copiotrophs or oligotrophs. Error bars represent 95% confidence intervals of the average value within the copiotroph ($n = 74$) and oligotroph ($n = 34$) categories. The copiotrophs were shown to have significantly more histidine kinases per gene than the oligotrophs by a Student's t test ($P = 3e^{-15}$).

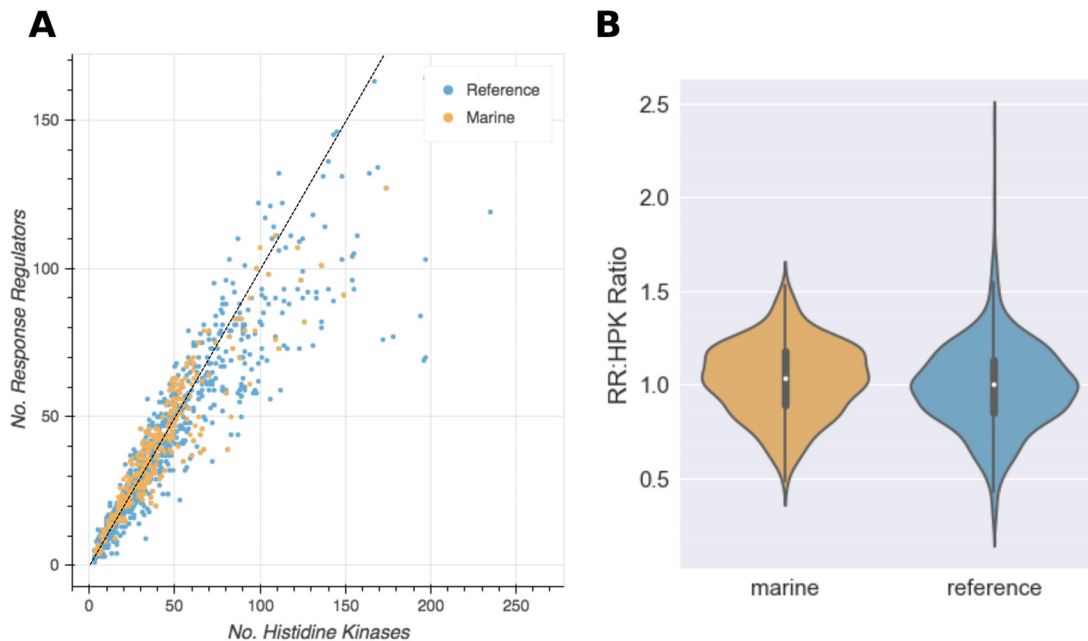


FIG 4 (A) Number of response regulator (RR) genes versus histidine kinase (HPK) genes in marine (orange) and reference (blue) bacteria. The dotted black line represents a 1:1 relationship. When the number of TCS systems is low, the ratio of RRs to HPKs is approximately 1. When the number of two-component systems is large (50 or more), the RR-to-HPK ratio tends to be much lower than 1. (B) RR/HPK ratio of marine and reference bacteria. While the differences are subtle, the marine bacteria have a significantly larger RR/HPK ratio on average (1.03) than the reference bacteria (0.99) as shown by a one-tailed ANOVA test [$P = 0.0095$, $F(327,1151) = 6.73$].

mental material). An important caveat is that both the marine and reference data sets contained mainly cultured organisms and may not represent natural diversity (13).

Based on the traditional understanding, the numbers of histidine kinase and response regulator genes are expected to be equal. Indeed, we found the average response regulator/histidine kinase (RR/HPK) ratio in the reference data set to be 0.99 (Fig. 4). The RR/HPK ratio was found to be slightly higher in the marine bacteria (1.03), suggesting that there are a small number of “extra” response regulators in the genomes. The difference between the marine and reference data sets is significant based on a one-way analysis of variance (ANOVA) test [$P = 0.009$, $F(372, 1151) = 6.73$].

To better understand the origin of the high RR/HPK ratio, we examined the locations of the TCS genes in a subselection of marine genomes (Fig. 5). Genomes were selected for their oceanographic relevance, quality, and representation in the literature. Typically, the TCS system histidine kinase and response regulator genes are located in the same operon (14). “Orphan” genes are defined here as genes that are more than four average gene lengths away from another TCS gene. These genes may participate in regulatory networks with other TCS systems (15).

Orphan TCS genes were identified in all seven of the genomes that we examined, consistent with the high RR/HPK ratios described above. For example, the *Pelagibacter ubique* HTCC1016 genome has three modular TCS systems plus one orphan HPK (a Kipl family gene) and one orphan RR (RegB). Notably, Kipl is known to participate in regulatory networks (15). Larger genomes have more orphan genes. For instance, *Alteromonas* sp. Alt199 has 26 orphan histidine kinases (30% of the histidine kinase genes) and 8 orphan response regulators (14% of the response regulator genes). *Crocospaera* sp. WH8501 has 8 orphan histidine kinases (15%) and 9 orphan response regulators (17%). It is difficult to ascertain the function of orphan genes due to the lack of experimental evidence associated with them. These systems are ideal for future study.

Hybrid systems occur when the histidine kinase and the response regulator are located on a single protein; here they were identified as genes containing both a

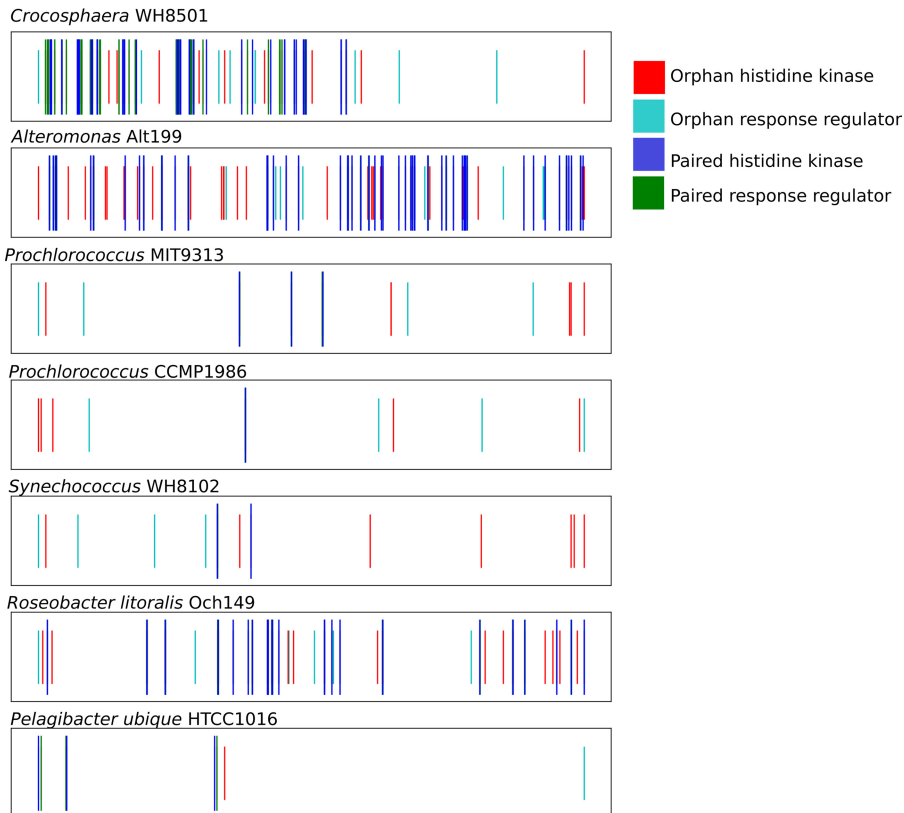


FIG 5 Distribution of TCS genes in various marine bacterial species. The genome is linearized and depicted as a number line; genes are represented as vertical lines based on their starting location. Histidine kinase genes and response regulator genes that are within four genes of another TCS gene are represented as long blue and green lines, respectively. Orphan histidine kinases and response regulators are represented as short cyan and red lines, respectively. There are many orphan TCS genes in marine bacteria, including in oligotrophs such as *Pelagibacter* and *Prochlorococcus*.

histidine kinase HAMP domain and a response regulator receiver domain. In marine bacteria, the proportion of hybrid histidine kinases relative to total histidine kinase content ranged from zero to 61% (Fig. 6). Marine bacteria have significantly more hybrid histidine kinases than reference bacteria ($P = 3e^{-10}$ by a Student's t test).

Patterns in *Proteobacteria* and *Cyanobacteria*. We examined the TCS systems of *Proteobacteria* and *Cyanobacteria* to explore differences among phylogenetically related organisms. *Proteobacteria* tend to have many histidine kinases (Fig. 7E). As before,

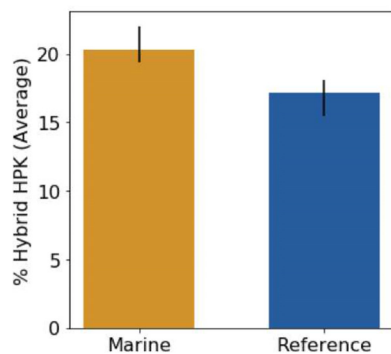


FIG 6 Percentage of histidine kinases that are hybrids in marine (orange) and reference (blue) bacteria. The marine bacteria have a greater percentage of hybrid histidine kinases than the reference bacteria. Error bars represent a bootstrapped 95% confidence interval. The difference is statistically significant by a Student's t test ($P = 3e^{-10}$).

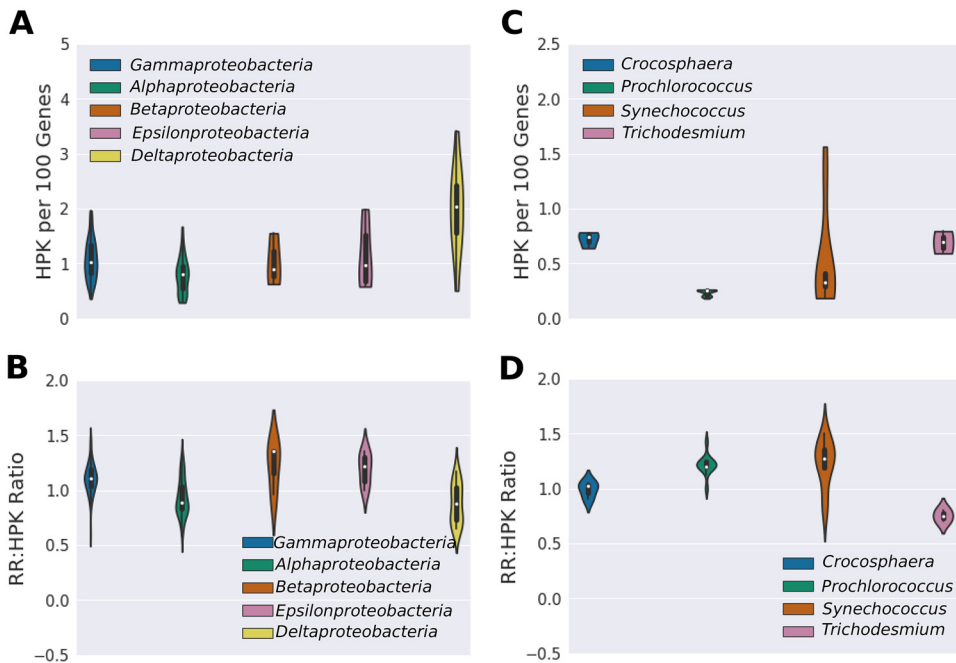


FIG 7 (A) Number of histidine kinases per 100 protein-encoding genes and (B) RR/HPK ratio of *Proteobacteria*. (C) Number of histidine kinases per 100 protein-encoding genes and (D) RR/HPK ratio of *Cyanobacteria*. *Proteobacteria*, in particular the *Deltaproteobacteria*, have many histidine kinases per gene compared with the *Cyanobacteria*. The RR/HPK ratio of the *Proteobacteria* tends to be greater than 1. Note that *Crocosphaera*, *Synechococcus*, and *Trichodesmium* have more histidine kinases per 100 protein-encoding genes than *Prochlorococcus*, which is adapted to highly oligotrophic conditions. The picocyanobacteria *Prochlorococcus* and *Synechococcus* have particularly high RR/HPK ratios.

we found that the *Deltaproteobacteria* devote a large portion of their genome to histidine kinases (see Fig. 3). In the *Cyanobacteria*, we observed more variation, with *Prochlorococcus* and *Synechococcus* tending to have few histidine kinases and the nitrogen-fixing diazotrophs having more (Fig. 7C).

Proteobacteria tend to have RR/HPK ratios greater than 1 (1.03 on average), suggesting the possibility of the presence of extra response regulators in the genomes. The RR/HPK ratios of the *Cyanobacteria* are more variable than those of the *Proteobacteria*. *Prochlorococcus* and *Synechococcus*, for instance, have very high RR/HPK ratios (1.22 and 1.23, respectively), indicating that there are many extra response regulators in the genome, while *Trichodesmium* organisms have a low RR/HPK ratio (0.75), indicating the possibility of the presence of extra histidine kinases.

Biogeochemical relevance of two-component sensory systems. To investigate whether TCS proteins can be used as biomarkers of oceanographic processes, we examined the distribution of a putative *Prochlorococcus* MIT9312 phosphate-sensing response regulator (PMT9312_0717) in metaproteomes of the tropical Pacific (Fig. 8). Data acquisition and analysis were previously described by Saito et al. in November 2011 (16). We identified two unique peptides from this protein. Using the open-source Metatryp software package (17), we determined that one peptide is specific to *Prochlorococcus* and that the other is specific to the order *Synechococcales*. Thus, the distribution presented here can be thought of as a general picocyanobacteria signal (Fig. 8D).

PMT9312_0717 was confined to the upper euphotic zone and was less prevalent in phosphate-rich waters near the equator (Fig. 8A). Notably, protein abundance was inversely correlated to phosphate concentration; the relationship can be modeled with a simple power law ($r^2 = 0.64$) (Fig. 8B and C). *Prochlorococcus* is highly abundant throughout the transect, while the protein is not, suggesting that PMT9312_0717 is specifically involved in phosphate regulation (Fig. S2). However, as with many orphan

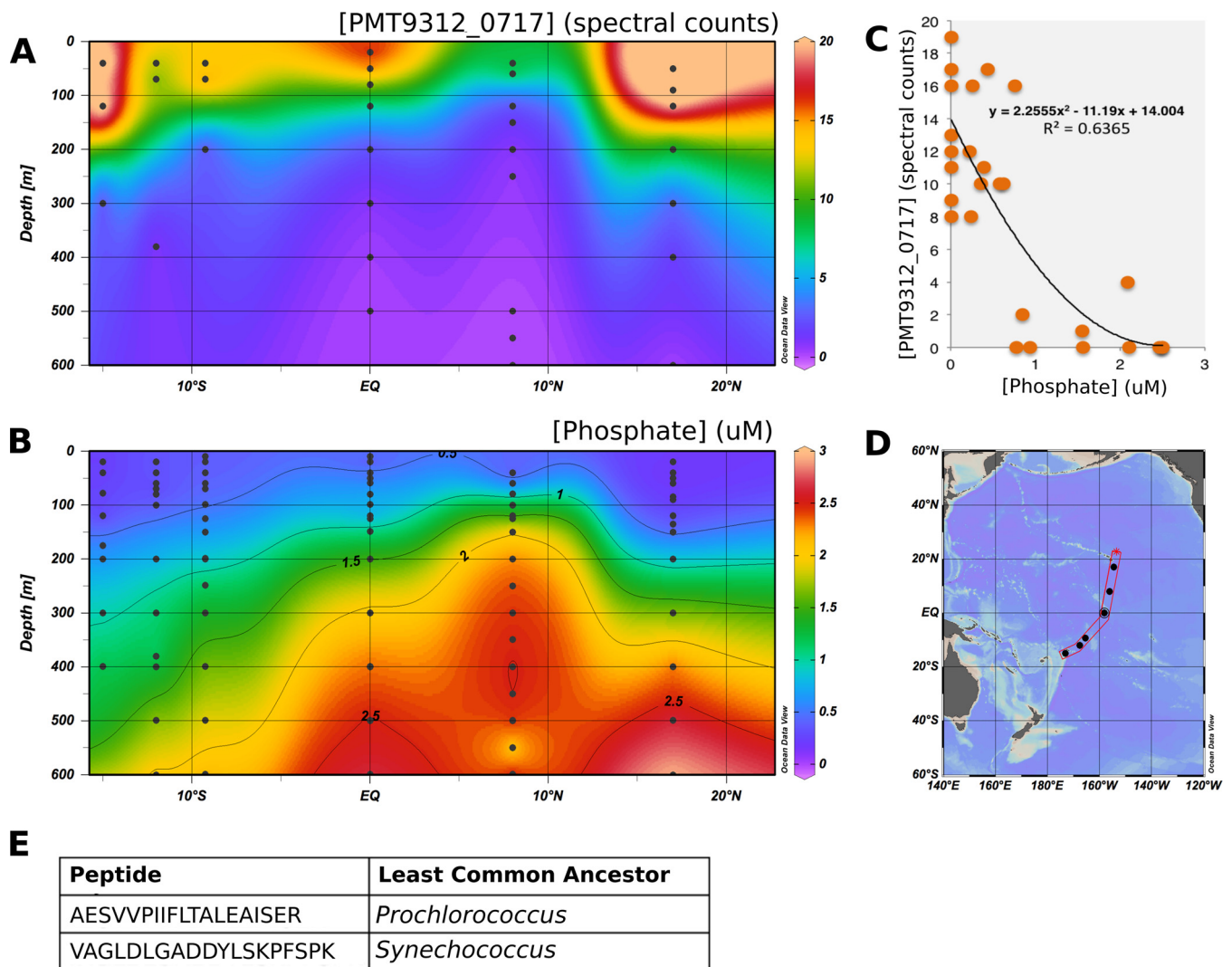


FIG 8 (A) Distribution of response regulator PMT9312_0717 in metaproteomes of the South Pacific Ocean. (B) Distribution of dissolved phosphate in the water column. (C) The abundance of PMT9312_0717 (measured as spectral counts) is correlated to phosphate concentrations. The relationship can be modeled by a power law. The distribution of *Prochlorococcus* cells is high across the transect (see Fig. S2). (D) Map of the METZYME transect where these samples were acquired. (E) Taxonomic information for the identified peptides. Two peptides were identified, one of which was annotated separately to two different *Prochlorococcus* strains. METATYP analysis suggests that both of these peptides are specific to the order *Synechococcales*.

TCS systems in marine bacteria, we were unable to identify a histidine kinase with corresponding relationships to phosphate concentrations. Thus, the exact regulatory function and mechanism of PMT9312_0717 remain a mystery, despite its probable biogeochemical relevance.

DISCUSSION

Two-component sensory systems allow bacteria to directly sense their internal and external surroundings and therefore play key roles in bacterial “intelligence” (18, 19). Because TCS systems are the most common regulatory systems in bacteria, studying them can provide insight into how individual cells might interact with their surroundings. As a result, TCS systems are of potentially significant ecological and biogeochemical importance, and yet they have been little studied in this context (7, 16). In this study, we surveyed the TCS systems of 328 diverse marine bacteria and described patterns in marine two-component sensing. We identified evidence of regulatory networking that may distinguish marine bacteria from other organisms and demonstrated that the abundance of a TCS system protein can be linked to oceanographic patterns.

Lifestyle influences TCS gene abundance. As in prior surveys, we found that the number of histidine kinases in the genome is largely determined by genome size (2, 4). However, the proportion of the genome encoding histidine kinases varies and is related to lifestyle and niche. The concept of copiotrophy versus oligotrophy is a common theme in microbial oceanography due to pronounced patterns in nutrient abundance versus scarcity, respectively, found in ocean environments (9–12). It is similar to K versus r selection in that oligotrophs, which are adapted to low-nutrient conditions, grow significantly slower than copiotrophs, which are adapted to high-nutrient conditions. Oligotrophs have smaller genomes and cell sizes that allow the organism to thrive in resource-limited environments (Table 1).

The small genomes of oligotrophs are thought to be the result of genome streamlining processes that are in turn driven by the need to conserve energetic and elemental resources (nitrogen and phosphorus in particular) (9, 11, 19). Our analysis suggests that the nutritional and energetic costs of maintaining TCS systems in oligotrophic environments outweigh the regulatory benefits. Lack of TCS genes has previously been observed in extremely oligotrophic organisms such as *Pelagibacter*, which may rely on simpler regulatory structures (such as one-component systems or riboswitches) which require fewer energetic and nutritional resources (20–22). The extent to which other marine oligotrophs may utilize simplified regulatory systems instead of or in addition to two-component systems is not yet known, but this survey found that the presence of few histidine kinases per gene is a hallmark of oligotrophy.

A previous effort to identify genomic signatures of oligotrophy found that histidine kinases were a weak indicator of copiotrophy versus oligotrophy (19). However, the study compared only two microbes, in contrast to the broader study across a larger number of diverse marine bacteria described here. The inability to adapt to rapid changes in nutrient availability and other environmental conditions may explain why oligotrophs are unable to survive in nutrient-rich environments (23). A recent comparison of coastal and open ocean strains of *Synechococcus* cyanobacteria supports this notion, where a coastal strain was found to have a more dynamic proteome response to iron scarcity than an oligotrophic strain (24).

In contrast to oligotrophs, certain organisms devote a comparatively large portion of their genome to histidine kinase genes. These bacteria often have specific traits such as mat formation that may require coordinated physiological alterations (25, 26). Many are sulfate or nitrate reducers from deep sea sediments or hydrothermal vents, which might represent particularly dynamic environments. Two-component systems may be particularly important for redox regulation in these organisms (27).

Unique patterns in marine TCS systems: RR/HPK ratios, orphan genes, and hybrid systems. We identified unique trends in the TCS system genes of marine bacteria. Specifically, we found that marine bacteria tend to have (i) higher RR/HPK ratios, (ii) many orphan TCS genes, and (iii) more hybrid histidine kinases than bacteria examined in a reference data set. We discuss each of these observations in turn.

Marine bacteria have significantly higher response regulator:histidine kinase ratios (RR:HPK) than bacteria as a whole, suggesting that they have extra response regulator genes. By considering the RR/HPK ratio, we found that approximately 1 in 50 response regulator genes lacks a histidine kinase partner (Fig. 4). The origin of the extra response regulators is puzzling. Some, lacking a histidine kinase partner, may have no regulatory function and could have been the result of incomplete genetic innovations or horizontal gene transfers (HGT). However, this explanation is not consistent with the tendency toward genome streamlining in the nutrient-limited ocean environment (9, 11). An intriguing alternative is that some of the extra response regulators might participate in regulatory networks in which multiple response regulators interact with a single histidine kinase. Such networks are a topic of increasing study and have been implicated in coordinating nutrient acquisition (specifically, acquisition of phosphate and iron), sporulation processes, stress response, and circadian rhythms (28–31).

Colocalization of TCS genes is thought to provide for concerted transcription of the sensory genes and better success in HGT (15, 32). Marine bacteria seem to be an exception to this rule, having many orphan genes in their genomes (Fig. 5). Orphan TCS genes are present in even the most streamlined genomes (i.e., that of *Pelagibacter ubique*), suggesting that they play important biochemical roles. The nonmodularity of TCS systems in marine bacteria suggests that the genes are not acquired through HGT but are instead acquired through gene duplication and genetic remodeling (32). TCS systems created in this way are thought to be more likely to participate in regulatory cross talk than systems that are acquired through horizontal gene transfer (3). Indeed, orphan genes are often involved in essential regulatory networks in model organisms (33–36). Alternatively, it is possible that in situations in which regulatory networks occur, the relationship between HPK and RR is not as specific as in normal two-component systems. This lack of specificity could allow nonmodular TCS genes to become fixed in the genome. Most of what we know about two-component systems is based on studies of modular systems from model organisms; additional studies on non-model organisms may thus reveal new mechanisms of acquisition and action of TCS genes.

In hybrid histidine kinases, the phosphorylation signal is relayed by one or more histidine phosphotransfer (Hpt) proteins before it reaches a terminal response regulator (Fig. 1). The added complexity of the phosphorelay is thought to provide multiple points of regulation, allowing fine-tuned physiological responses. For example, a hybrid histidine kinase regulates glycan utilization in *Bacteroides thetaiotaomicron* by integrating both intracellular metabolism and extracellular substrate signals (37). Hybrid histidine kinases are associated with physiological and behavioral complexity, being especially prevalent in higher eukaryotes (38). Marine bacteria such as *Proteobacteria* and *Bacteroidetes* can have many hybrid TCS systems (Fig. 5; see also Table 1), concurrent with a tendency toward metabolic complexity and particle association, which may drive multicellular behaviors (28).

Together, the presence of extra response regulators, the nonmodularity of TCS systems, and the prevalence of hybrid TCS systems suggest increased regulatory complexity in marine bacteria. This may confer advantages in the ocean environment, where bacteria are often chronically nutrient limited. In the ocean, nutrient availability is determined by diffusion rates, resulting in covariance in the distribution of organic nitrogen, carbon, phosphorus, and trace nutrients, especially at the microscale (28, 39). Nutrient colimitation has been demonstrated in a number of marine environments and may be more prevalent than originally thought (40–42). Marine organisms appear to have specific physiological responses to colimitation, for example, proteome restructuring and cell size decreases in iron and phosphate colimited *Trichodesmium* cells (43). Although the regulatory systems for colimitation in marine microbes have yet to be elucidated, a precedent for regulatory networking has been identified in nonmarine organisms such as *Edwardsiella tarda*, in which the Pho and Fur systems interact with one another (31). The irregularities in marine bacterial TCS genes suggest that similar regulatory networks may underpin concerted responses to multiple environmental perturbations.

Comparison of *Proteobacteria* and *Cyanobacteria*. *Proteobacteria* and *Cyanobacteria* are perhaps the most abundant and well-studied cells in the ocean. Comparing them demonstrates the impact of lifestyle traits on the number of histidine kinases in the genome. With the exception of the oligotrophic *Pelagibacter* species, the *Proteobacteria* have a larger proportion of genes encoding histidine kinases (0.95 per 100 protein-encoding genes on average) than the *Cyanobacteria* (0.57 per 100 protein-encoding genes on average) (Fig. 7A and C). This may be related to their tendency toward copiotrophy, which is associated with large numbers of TCS system genes. Within the *Cyanobacteria*, nitrogen-fixing organisms have more histidine kinases. Diazotrophs are not subjected to the same genome streamlining pressure as other marine *Cyanobacteria* owing to the fact that they can access an unlimited supply of atmo-

spheric nitrogen. However, the complexities of the diazotrophic lifestyle may necessitate a large number of regulatory genes. For instance, nitrogen fixation rates are known to respond to many environmental parameters such as iron nitrogen, phosphorus, dust, and light availability (44–48). Reflecting this, marine diazotrophs have two-component systems regulating nutrient availability, complex circadian rhythms, and redox state.

Evidence for regulatory networking is prevalent in both the *Proteobacteria* and *Cyanobacteria* (Fig. 7B and D). Most *Proteobacteria* and picocyanobacteria have elevated RR/HPK ratios that suggest the presence of branched regulatory networks in which one histidine kinase communicates with multiple response regulators. Branched regulatory networks could allow multiple operons to be affected by a single sensory input, with each chemical sensor triggering multiple downstream effects, providing greater metabolic flexibility and dynamism. This could provide for fine-tuned responses to multiple stimuli, such as nutrient colimitation. In nutrient-limited environments, regulatory networking may provide an advantage for cellular resource conservation. For instance, a single stimulus can trigger multiple physiological reactions without the need to express an entire two-component system for each operon. Consistently, we found that the oligotrophic genomes from organisms such as *Pelagibacter* and *Prochlorococcus* have especially high RR/HPK ratios (Table 1).

Notably, the diazotrophic organism *Trichodesmium* has low RR/HPK ratios, suggesting regulatory networking in which multiple histidine kinases interact with a single response regulator. This may allow integration of multiple environmental signals in regulating a single physiological response and may underpin extensive proteomic changes such as coordination of carbon and nitrogen fixation processes over the course of the diel cycle (49). However, identifying these networks is challenging because the majority of two-component-system genes in *Trichodesmium* (and many marine bacteria) have not been characterized.

Two-component systems as potential biogeochemical biomarkers. Having so far considered TCS genes, we next turned to the gene products and their relationship to oceanographic processes. We studied a *Prochlorococcus/Synechococcus* PhoB-like response regulator protein, PMT9312_0717, in the tropical Pacific Ocean. PMT9312_0717 is one of the most abundant TCS system proteins in this transect. It is an orphan, and its partner histidine kinase is not known. In the metaproteomics analysis, we found that the protein abundance is inversely correlated with inorganic phosphate concentration, particularly at concentrations below 1 μM phosphate such as are encountered near the nutricline (Fig. 8). The abundance of *Prochlorococcus* cells is high throughout the transect, while that of PMT9312_0717 is not and instead follows trends in phosphate concentrations (see Fig. S2 in the supplemental material). This implies that the abundance of the protein is related to oceanographic processes, suggesting that the protein self-regulates its production. Increased abundance of TCS systems in response to low dissolved phosphorus concentrations has been previously observed for phosphate-regulating systems (7, 16). Measurement of levels of protein phosphorylation (i.e., the activity of the TCS system), while technically difficult, could provide additional information about the function of the protein.

In addition to this protein distribution, genomic evidence also suggests that PMT9312_0717 is involved in phosphate regulation. PMT9312_0717 is present in many *Prochlorococcus* species, including both high- and low-light ecotypes, as well as in *Synechococcus* species. It is similar to both the *Prochlorococcus* sp. 9312 phosphate-sensing histidine kinase PhoB (37% identity) and the analogous *Synechococcus* sp. WH8102 PhoP protein (WP_025362545.1; 37% identity) (16) but is a distinct protein. Because just a few point mutations can change the function of a response regulator, we hypothesized that this protein participates in phosphate regulation but in ways that differ from those seen with PhoB/PhoP. Corroborating its possible role in phosphate sensing, PMT9312_0717 is located near a DedA family alkaline phosphatase-related gene (e.g., PMT9312_0712) in multiple strains of *Prochlorococcus*. TCS systems have amino acids known as specificity residues that govern the histidine kinase-response

regulator interaction (4). The specificity residues of the PMT9312_0717 and PhoB genes in *Prochlorococcus* sp. MIT 9312 are not shared, indicating that the PMT9312_0717 is unlikely to interact with the phosphate-sensing PhoR histidine kinase.

TCS systems underpin many of the physiological changes observed in laboratory and field perturbations of marine bacteria. They are involved in nutrient acquisition, detoxification, quorum sensing, and other topical themes in marine microbiology. However, our knowledge of these systems in marine species is limited. Sequence-based identification of TCS systems is difficult because histidine kinases and response regulators share highly conserved catalytic domains. Thus, while it is possible to identify TCS genes, identifying their physiological functions is tricky. Few two-component systems have been experimentally verified in marine species, despite the fact that just a few amino acid substitutions can drastically change physiological function (4).

The environmental drivers behind gain/loss of two-component sensory system genes and protein synthesis processes are not well understood. For instance, distribution of the phosphate-sensing *phoR-phoB* two-component genes in *Prochlorococcus*, while initially hypothesized to be correlated to phosphate availability, cannot be consistently linked to large-scale oceanographic patterns (7, 50). An important consideration is that two-component sensory systems act on short time scales—seconds, minutes, or hours (51, 52). The presence of a TCS system thus suggests the need to continuously monitor the stimulus. For nutrient-sensing regulators, it may be more accurate to suggest that distribution of the TCS system is related to varying (not necessarily chronically depleted) nutrient concentrations such as are found in surface waters. This is corroborated by our finding that the amount of PMT9312_0717 protein in the water column increases significantly near nutricline like concentrations. Given this circumstantial and yet consistent evidence, a detailed biochemical characterization of PMT9312_0717 is an intriguing topic for future study. However, on the basis of this and previous work, it is clear that TCS system protein abundances can contain valuable biogeochemical information (15).

Conclusions. Two-component sensory systems reveal characteristics of both individual cells and the ecosystems in which they live. In this way, they represent a unique opportunity to link microbial physiology to the environment. We know relatively little about TCS systems in marine bacteria, but it is clear that the distribution of TCS genes and proteins is dependent on both the traits of the organism and its surrounding environment. For instance, we found that oligotrophs have significantly fewer histidine kinases per gene than copiotrophs and that diazotrophy is associated with greater numbers of TCS system genes. Importantly, we found that marine microbes may have adapted unique ways to sense their environments using complex regulatory networks. Additional characterization of these networks may provide us with a greater appreciation for both the uniqueness of the ocean environment and the breadth of sensory systems used by prokaryotes. Detailed biochemical characterization of marine two-component systems is greatly needed and has great potential to advance our understanding of microbial life and its connections to global biogeochemical cycles.

MATERIALS AND METHODS

Marine and reference bacteria data sets. We acquired the genomes of 328 marine bacteria available in the JGI IMG data warehouse (see Table S1 in the supplemental material). All of the genomes were high-quality finished genomes or permanent drafts. We note that permanent draft genomes could be missing some RR or HPK genes due to incomplete assembly, but this is not expected to significantly affect results. The marine data set is phylogenetically diverse; the best represented phyla are the *Cyanobacteria*, *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, and *Actinobacteria* (see Fig. S1 in the supplemental material). The bacteria were isolated from various habitats that included coastal ecosystems, the open ocean, hydrothermal vent systems, marine sediments, and microbial mats. Most of the genomes that we examined were from bacterial isolates in culture; as such, bacteria of particular oceanographic interest such as *Prochlorococcus*, as well as easily cultivated organisms such as *Alteromonas*, form a large fraction of the data set. While we attempted to capture maximal phylogenetic and habitat level variability, this data set does not necessarily represent the actual levels of bacterial diversity nor cell abundance in the ocean environment. Rather, the intention was to allow us to examine the marine bacteria for which we have good-quality genomic information at this time.

The reference bacteria data set was largely derived from the organisms associated with the GEBA-I initiative, which provides a collection of bacterial genomes spanning the breadth of known phylogenetic diversity (Table S2) (8). The GEBA-I database includes genomes of both bacteria and archaea; we used only the bacterial genomes here. We observed that *Cyanobacteria* are underrepresented in the GEBA-I data set relative to the marine data set described above. To facilitate comparisons between the marine and reference data sets, we included 180 high-quality, nonmarine *Cyanobacteria* genomes from the JGI IMG warehouse in the reference data set.

Identification of two-component-system genes. We identified TCS system genes by protein family (pfam) domain annotations (53). The pfam domain annotation was performed for each genome as part of their initial ingestion into the IMG warehouse. Comparisons were thus facilitated by the fact that all genomes were subjected to similar annotation analysis pipelines in the JGI IMG portal.

Histidine kinases are composed of two domains—an HATPase domain and a phosphoacceptor domain—which represent separate entries in the pfam domain database. We tested both the HATPase and phosphoacceptor domains by comparing the number of histidine kinases identified with the results of an extensively curated previously published survey of TCS genes in bacteria (18). We found that the HATPase domain was better conserved, identified more histidine kinases, and resulted in histidine kinase identifications comparable to those in the previous study. Thus, we used the HATPase domain for histidine kinase identification subsequently. The specific pfams used were pfam02518 (HATPase_c), pfam13581 (HATPase_c_2), pfam13589 (HATPase_c_3), pfam14501 (HATPase_c_5), and pfam07536 (HWE_HK). The HATPase domain is also present in proteins such as DNA gyrase (pfam00204), HSP90 (pfam00183), and MutL (pfam13941); we removed genes containing these three pfams from the analysis.

Response regulators are composed of a phosphoreceiver domain (sometimes known as a REC domain) and an output domain. The output domain can vary significantly and determines the biological function of the response regulator (52, 54). We used the protein family domain for the response regulator receiver (pfam00072) to identify response regulators in our data sets. When possible, we compared the results of this pfam-based identification of response regulators to that of a previously published survey and found that the results of the pfam-based analysis were comparable (18).

Hybrid histidine kinases occur when the response regulator phosphoacceptor domain is present on the same protein as the histidine kinase sensory and phosphoacceptor domains. We define them here as genes containing both a HATPase domain (pfam02518, pfam13581, pfam13589, pfam14501, or pfam07730) and the response regulator phosphoreceiver domain (pfam00072). Thus, the hybrid histidine kinases are present in both the histidine kinase data and response regulator data for a given genome.

Statistical and meta-analyses. All analyses were conducted in Python 3.6. The entire data analysis and visualization pipeline, including statistical analyses and machine learning algorithms, can be recreated by accessing the scripts at https://github.com/naheld/patterns_TCS_sensing_marine_bacteria. A fully executable cloud environment is provided courtesy of the Binder project (<https://mybinder.org/>). We used the Bokeh (<https://bokeh.pydata.org/en/latest/>) and seaborn (<https://seaborn.pydata.org/>) libraries to generate visualizations and perform statistical analyses. For ratio and statistical analyses, genomes containing 0 histidine kinases and/or 0 response regulators were excluded.

For K means clustering analyses, we counted the number of histidine kinases, including hybrid genes, and normalized the data by dividing by the number of protein-encoding genes. For human readability, we express this value as the number of histidine kinases per 100 genes. To eliminate the effect of scaling, we subjected the data to unit normalization. We then performed the clustering analysis to identify groups of genomes with similar signaling repertoires in relation to genome size. We used the Silhouette method implemented in the Scikit Learn Cluster module to select the optimal number of clusters (4) by maximizing the average Silhouette coefficient values (55). Clustering analysis was performed with the Scikit Learn K-means clustering algorithm.

Locations of TCS genes. The locations of TCS genes in the genomes of marine bacteria were examined by plotting the gene starting locations on a linearized depiction of the circular bacterial genome (a number line). The average gene size was calculated for each genome examined. A gene was considered to be an orphan if its start point was farther than four average gene lengths from another TCS gene.

PhoB protein distribution analysis. We examined the distribution of response regulator PMT9312_0717 in marine metaproteomes from the METZYME expedition in the tropical Pacific. Analysis of these metaproteomes has been previously described (16). Briefly, microbial biomass was collected with *in situ* particle collection pumps (McLane Labs) on the KM1128 METZYME research expedition. Proteins from the 0.2- μ M to 3- μ M size fraction were subjected to SDS detergent extraction and in-gel trypsin digestion as previously described (16). Peptides were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using 1-dimensional chromatography on a Thermo Fusion Orbitrap mass spectrometer (16). Spectral counts for the protein were generated by mapping against 6 metagenomes sampled from the METZYME expedition. These were sequenced at JGI and assembled using metaSPAdes (56). Genes were predicted and annotated using the pipeline described previously by Dupont et al. (57). Peptide-to-spectrum matches (PSMs) were identified by SEQUEST and restricted to a 10-ppm peptide mass threshold and a 99.0% protein probability threshold (16). Two unique peptides were identified for PMT9312_0717; we performed redundancy analysis using the openly available Metatryp software (17). Corresponding inorganic phosphate analyses were conducted by Joe Jennings at Oregon State University as previously described (58).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00317-18>.

FIG S1, EPS file, 0.1 MB.

FIG S2, EPS file, 1 MB.

TABLE S1, XLSX file, 0.2 MB.

TABLE S2, XLSX file, 0.2 MB.

ACKNOWLEDGMENTS

We thank Joe Jennings at Oregon State University and Chris Dupont at the J. Craig Venter Institute for providing nutrient and metagenomic analyses, respectively, for the KM1128 METZYME research expedition. We also thank our anonymous reviewers for their thoughtful comments.

This material is based on work supported by a National Science Foundation Graduate Research Fellowship under grant number 1122274 (N. A. Held). It was also supported by the Gordon and Betty Moore Foundation (grant number 3782 [M. Saito]) and by the National Science Foundation (grant numbers OCE-1657766, EarthCube 1639714, OCE-1658030, and OCE-1260233).

REFERENCES

- Fuhrman JA, Cram JA, Needham DM. 2015. Marine microbial community dynamics and their ecological interpretation. *Nat Rev Microbiol* 13: 133–146. <https://doi.org/10.1038/nrmicro3417>.
- Galperin MY, Higdon R, Kolker E. 2010. Interplay of heritage and habitat in the distribution of bacterial signal transduction systems. *Mol Biosyst* 6:721–728. <https://doi.org/10.1039/b908047c>.
- Heermann R, Jung K. 2010. Stimulus perception and signaling in histidine kinases, p 135–161. *In* Krämer R, Jung K (ed), *Bacterial signaling*. John Wiley & Sons, Inc, Hoboken, NJ. <https://doi.org/10.1002/9783527629237.ch8>.
- Capra EJ, Laub MT. 2012. Evolution of two-component signal transduction systems. *Annu Rev Microbiol* 66:325–347. <https://doi.org/10.1146/annurev-micro-092611-150039>.
- Gao R, Mack TR, Stock AM. 2007. Bacterial response regulators: versatile regulatory strategies from common domains. *Trends Biochem Sci* 32: 225–234. <https://doi.org/10.1016/j.tibs.2007.03.002>.
- Yamada M, Makino K, Amemura M, Shinagawa H, Nakata A. 1989. Regulation of the phosphate regulon of *Escherichia coli*: analysis of mutant *phoB* and *phoR* genes causing different phenotypes. *J Bacteriol* 171:5601–5606. <https://doi.org/10.1128/jb.171.10.5601-5606.1989>.
- Martiny AC, Coleman ML, Chisholm SW. 2006. Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc Natl Acad Sci U S A* 103:12552–12557. <https://doi.org/10.1073/pnas.0601301103>.
- Mukherjee S, Seshadri R, Varghese NJ, Eloë-Fadrosch EA, Meier-Kolthoff JP, Göker M, Coates RC, Hadjithomas M, Pavlopoulos GA, Paez-Espino D, Yoshikuni Y, Visel A, Whitman WB, Garrity GM, Eisen JA, Hugenholtz P, Pati A, Ivanova NN, Woyke T, Klenk HP, Kyrpides NC. 2017. 1003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat Biotechnol* 35:676–683. <https://doi.org/10.1038/nbt.3886>.
- Swan BK, Tupper B, Sczyrba A, Lauro FM, Martínez-García M, González JM, Luo H, Wright JJ, Landry ZC, Hanson NW, Thompson BP, Poulton NJ, Schwientek P, Acinas SG, Giovannoni SJ, Moran MA, Hallam SJ, Cavicchioli R, Woyke T, Stepanauskas R. 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A* 110:11463–11468. <https://doi.org/10.1073/pnas.1304246110>.
- Kirchman DL. 2016. Growth rates of microbes in the oceans. *Annu Rev Mar Sci* 8:285–309. <https://doi.org/10.1146/annurev-marine-122414-033938>.
- Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ, Rappé MS. 2012. Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *mBio* 3:e00252-12. <https://doi.org/10.1128/mBio.00252-12>.
- Ivars-Martinez E, Martin-Cuadrado A-B, D'Auria G, Mira A, Ferreria S, Johnson J, Friedman R, Rodríguez-Valera F. 2008. Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter. *ISME J* 2:1194–1212. <https://doi.org/10.1038/ismej.2008.74>.
- Staley JT, Konopka A. 1985. Measurement of in situ activities of non-photosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* 39:321–346. <https://doi.org/10.1146/annurev.mi.39.100185.001541>.
- Barakat M, Ortet P, Whitworth DE. 2011. P2CS: a database of prokaryotic two-component systems. *Nucleic Acids Res* 39:771–776. <https://doi.org/10.1093/nar/gkq1023>.
- Mitrophanov AY, Groisman EA. 2008. Signal integration in bacterial two-component regulatory systems. *Genes Dev* 22:2601–2611. <https://doi.org/10.1101/gad.1700308>.
- Saito MA, Mcllvain MR, Moran DM, Goepfert TJ, DiTullio GR, Post AF, Lamborg CH. 2014. Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science* 345:1173–1177. <https://doi.org/10.1126/science.1256450>.
- Saito MA, Dorsk A, Post AF, Mcllvain MR, Rappé MS, DiTullio GR, Moran DM. 22 July 2015. Needles in the blue sea: sub-species specificity in targeted protein biomarker analyses within the vast oceanic microbial metaproteome. *Proteomics* <https://doi.org/10.1002/pmic.201400630>.
- Galperin MY. 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol* 5:35. <https://doi.org/10.1186/1471-2180-5-35>.
- Lyon P. 2015. The cognitive cell: bacterial behavior reconsidered. *Front Microbiol* 6:264–218. <https://doi.org/10.3389/fmicb.2015.00264>.
- Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S, DeMaere MZ, Ting L, Ertan H, Johnson J, Ferreria S, Lapidus A, Anderson I, Kyrpides N, Munk AC, Detter C, Han CS, Brown MV, Robb FT, Kjelleberg S, Cavicchioli R. 2009. The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci U S A* 106:15527–15533. <https://doi.org/10.1073/pnas.0903507106>.
- Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, Bibbs L, Eads J, Richardson TH, Noordewier M, Rappé MS, Short JM, Carrington JC, Mathur EJ. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309:1242–1245.
- Meyer MM, Ames TD, Smith DP, Weinberg Z, Schwalbach MS, Giovannoni SJ, Breaker RR. 2009. Identification of candidate structured RNAs in the marine organism ‘*Candidatus Pelagibacter ubique*’. *BMC Genomics* 10:116. <https://doi.org/10.1186/1471-2164-10-268>.
- Koch AL. 2001. Oligotrophs versus copiotrophs. *BioEssays* 23:657–661. <https://doi.org/10.1002/bies.1091>.
- Mackey KRM, Post AF, Mcllvain MR, Cutter GA, John SG, Saito MA. 2015. Divergent responses of Atlantic coastal and oceanic *Synechococcus* to

- iron limitation. *Proc Natl Acad Sci U S A* 112:9944–9949. <https://doi.org/10.1073/pnas.1509448112>.
25. Lehti TA, Heikkinen J, Korhonen TK, Westerlund-Wikstrom B. 2012. The response regulator RcsB activates expression of Mat fimbriae in meningitic *Escherichia coli*. *J Bacteriol* 194:3475–3485. <https://doi.org/10.1128/JB.06596-11>.
 26. Reynolds TB, Jansen A, Peng X, Fink GR. 2008. Mat formation in *Saccharomyces cerevisiae* requires nutrient and pH gradients. *Eukaryot Cell* 7:122–130. <https://doi.org/10.1128/EC.00310-06>.
 27. Kusian B, Bowien B. 1997. Organization and regulation of *cbb* CO₂ assimilation genes in autotrophic bacteria. *FEMS Microbiol Rev* 21: 135–155. <https://doi.org/10.1111/j.1574-6976.1997.tb00348.x>.
 28. Lyons NA, Kolter R. 2015. On the evolution of bacterial multicellularity. *Curr Opin Microbiol* 24:21–28. <https://doi.org/10.1016/j.mib.2014.12.007>.
 29. Espinosa J, Boyd JS, Cantos R, Salinas P, Golden SS, Contreras A. 2015. Cross-talk and regulatory interactions between the essential response regulator RpaB and cyanobacterial circadian clock output. *Proc Natl Acad Sci U S A* 112:2198–2203. <https://doi.org/10.1073/pnas.1424632112>.
 30. Chakraborty S, Sivaraman J, Leung KY, Mok YK. 2011. Two-component PhoB-PhoR regulatory system and ferric uptake regulator sense phosphate and iron to control virulence genes in type III and VI secretion systems of *Edwardsiella tarda*. *J Biol Chem* 286:39417–39430. <https://doi.org/10.1074/jbc.M111.295188>.
 31. Yamamoto K, Hirao K, Oshima T, Aiba H, Utsumi R, Ishihama A. 2005. Functional characterization in vitro of all two-component signal transduction systems from *Escherichia coli*. *J Biol Chem* 280:1448–1456. <https://doi.org/10.1074/jbc.M410104200>.
 32. Alm E, Huang K, Arkin A. 2006. The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. *PLoS Comput Biol* 2:1329–1342. <https://doi.org/10.1371/journal.pcbi.0020143>.
 33. Overgaard M, Wegener-Feldbrugge S, Sogaard-Andersen L. 2006. The orphan response regulator DigR is required for synthesis of extracellular matrix fibrils in *Myxococcus xanthus*. *J Bacteriol* 188:4384–4394. <https://doi.org/10.1128/JB.00189-06>.
 34. Idone V, Brendtro S, Gillespie R, Kocaj S, Peterson E, Rendi M, Warren W, Michalek S, Krastel K, Cvitkovitch D, Spatafora G. 2003. Effect of an orphan response regulator on *Streptococcus mutans* sucrose-dependent adherence and cariogenesis. *Infect Immun* 71:4351–4360. <https://doi.org/10.1128/IAI.71.8.4351-4360.2003>.
 35. Lu Y, He J, Zhu H, Yu Z, Wang R, Chen Y, Dang F, Zhang W, Yang S, Jiang W. 2011. An orphan histidine kinase, OhkA, regulates both secondary metabolism and morphological differentiation in *Streptomyces coelicolor*. *J Bacteriol* 193:3020–3032. <https://doi.org/10.1128/JB.00017-11>.
 36. Workentine ML, Chang L, Ceri H, Turner RJ. 2009. The GacS-GacA two-component regulatory system of *Pseudomonas fluorescens*: a bacterial two-hybrid analysis. *FEMS Microbiol Lett* 292:50–56. <https://doi.org/10.1111/j.1574-6968.2008.01445.x>.
 37. Sonnenburg ED, Sonnenburg JL, Manchester JK, Hansen EE, Chiang HC, Gordon JI. 2006. A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. *Proc Natl Acad Sci U S A* 103:8834–8839. <https://doi.org/10.1073/pnas.0603249103>.
 38. Wuichet K, Cantwell BJ, Zhulin IB. 2010. Evolution and phyletic distribution of two-component signal transduction systems. *Curr Opin Microbiol* 13:219–225. <https://doi.org/10.1016/j.mib.2009.12.011>.
 39. Stocker R. 2012. Marine microbes see a sea of gradients. *Science* 338: 628–633. <https://doi.org/10.1126/science.1208929>.
 40. Moore CM, Mills MM, Arrigo KR, Berman-Frank I, Bopp L, Boyd PW, Galbraith ED, Geider RJ, Guieu C, Jaccard SL, Jickells TD, La Roche J, Lenton TM, Mahowald NM, Marañón E, Marinov I, Moore JK, Nakatsuka T, Oschlies A, Saito MA, Thingstad TF, Tsuda A, Ulloa O. 2013. Processes and patterns of oceanic nutrient limitation. *Nature Geosci* 6:701–710. <https://doi.org/10.1038/ngeo1765>.
 41. Saito MA, Goepfert TJ, Ritt JT. 2008. Some thoughts on the concept of co-limitation: three definitions and the importance of bioavailability. *Limnol Oceanogr* 53:276–290. <https://doi.org/10.4319/lo.2008.53.1.0276>.
 42. Browning TJ, Eric P, Rapp I, Engel A, Bertrand EM, Tagliabue A, Moore CM. 2017. Nutrient co-limitation at the boundary of an oceanic gyre. *Nature* 551:242–246. <https://doi.org/10.1038/nature24063>.
 43. Walworth NG, Fu FX, Webb EA, Saito MA, Moran D, McIlvin MR, Lee MD, Hutchins DA. 2016. Mechanisms of increased *Trichodesmium* fitness under iron and phosphorus co-limitation in the present and future ocean. *Nat Commun* 7:1–11. <https://doi.org/10.1038/ncomms12081>.
 44. Karl D, Michaels A, Bergman B, Capone D, Carpenter E, Letelier R, Lipschultz F, Paerl H, Sigman D, Stal L. 2002. Dinitrogen fixation in the world's oceans. *Biogeochemistry* 57:47–98. <https://doi.org/10.1023/A:1015798105851>.
 45. Sanudo-Wilhelmy SA, Kustka AB, Gobler CJ, Hutchins DA, Yang M, Lwiza K, Burns J, Capone DG, Raven JA, Carpenter EJ. 2001. Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature* 411:66–69. <https://doi.org/10.1038/35075041>.
 46. Orchard ED, Webb EA, Dyrman ST. 2009. Molecular analysis of the phosphorus starvation response in *Trichodesmium* spp. *Environ Microbiol* 11:2400–2411. <https://doi.org/10.1111/j.1462-2920.2009.01968.x>.
 47. Chen YB, Chen YB, Dominic B, Mellon MT, Zehr JP. 1998. Circadian rhythm of nitrogenase gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium *Trichodesmium* sp. strain IMS101. *J Bacteriol* 180:3598–3605.
 48. Rubin M, Berman-Frank I, Shaked Y. 2011. Dust- and mineral-iron utilization by the marine dinitrogen-fixer *Trichodesmium*. *Nature Geosci* 4:529–534. <https://doi.org/10.1038/ngeo1181>.
 49. Saito MA, Bertrand EM, Dutkiewicz S, Bulygin VV, Moran DM, Monteiro FM, Follows MJ, Valois FW, Waterbury JB. 2011. Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph *Crocospheera watsonii*. *Proc Natl Acad Sci U S A* 108:2184–2189. <https://doi.org/10.1073/pnas.1006943108>.
 50. Luo C, Konstantinidis KT. 7 April 2011. Phosphorus-related gene content is similar in *Prochlorococcus* populations from the North Pacific and North Atlantic Oceans. *Proc Natl Acad Sci U S A* <https://doi.org/10.1073/pnas.1018662108>.
 51. Mary I, Vault D. 2003. Two-component systems in *Prochlorococcus* MED4: genomic analysis and differential expression under stress. *FEMS Microbiol Lett* 226:135–144. [https://doi.org/10.1016/S0378-1097\(03\)00587-1](https://doi.org/10.1016/S0378-1097(03)00587-1).
 52. Ogawa T, Bao DH, Katoh H, Shibata M, Pakrasi HB, Bhattacharyya-Pakrasi M. 2002. A two-component signal transduction pathway regulates manganese homeostasis in *Synechocystis* 6803, a photosynthetic organism. *J Biol Chem* 277:28981–28986. <https://doi.org/10.1074/jbc.M204175200>.
 53. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heeger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42:222–230. <https://doi.org/10.1093/nar/gkt1223>.
 54. Galperin MY. 2006. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J Bacteriol* 188:4169–4182. <https://doi.org/10.1128/JB.01887-05>.
 55. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. 2012. Scikit-learn: machine learning in Python. *J Machine Learning Res* 12:2825–2830. <https://doi.org/10.1007/s13398-014-0173-7.2>.
 56. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>.
 57. Dupont CL, McCrow JP, Valas R, Moustafa A, Walworth N, Goodenough U, Roth R, Hogle SL, Bai J, Johnson ZI, Mann E, Palenik B, Barbeau KA, Venter JC, Allen AE. 2015. Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *ISME J* 9:1076–1092. <https://doi.org/10.1038/ismej.2014.198>.
 58. Noble AE, Lamborg CH, Ohnemus DC, Lam PJ, Goepfert TJ, Measures CI, Frame CH, Casciotti KL, DiTullio GR, Jennings J, Saito MA. 2012. Basin-scale inputs of cobalt, iron, and manganese from the Benguela-Angola front into the South Atlantic Ocean. *Limnol Oceanogr* 57:989–1010. <https://doi.org/10.4319/lo.2012.57.4.0989>.
 59. Rappé MS, Connon SA, Vergin KL, Giovannoni SJ. 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* 418:630–633. <https://doi.org/10.1038/nature00917>.
 60. Moore LR, Goericke R, Chisholm SW. 1995. Comparative physiology of *Synechococcus* and *Prochlorococcus* - influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Mar Ecol Prog Ser* 116:259–275. <https://doi.org/10.3354/meps116259>.
 61. Hutchins DA, Fu F-X, Zhang Y, Warner ME, Feng Y, Portune K, Bernhard PW, Mulholland MR. 2007. CO₂ control of *Trichodesmium* N₂ fixation, photosynthesis, growth rates, and elemental ratios: implications for past, present, and future ocean biogeochemistry. *Limnol Oceanogr* 52: 1293–1304. <https://doi.org/10.4319/lo.2007.52.4.1293>.
 62. Webb EA, Ehrenreich IM, Brown SL, Valois FW, Waterbury JB. 2009.

- Phenotypic and genotypic characterization of multiple strains of the diazotrophic cyanobacterium, *Crocospaera watsonii*, isolated from the open ocean. *Environ Microbiol* 11:338–348. <https://doi.org/10.1111/j.1462-2920.2008.01771.x>.
63. Teira E, Gasol JM, Aranguren-Gassis M, Fernández A, González J, Lekunberri I, Álvarez-Salgado XA. 2008. Linkages between bacterioplankton community composition, heterotrophic carbon cycling and environmental conditions in a highly dynamic coastal ecosystem. *Environ Microbiol* 10:906–917. <https://doi.org/10.1111/j.1462-2920.2007.01509.x>.
 64. Mourin RR, Worden AZ, Azam F. 2003. Growth of *Vibrio cholerae* O1 in red tide waters off California. *Appl Environ Microbiol* 69:6923–6931. <https://doi.org/10.1128/AEM.69.11.6923>.
 65. López-Pérez M, Gonzaga A, Ivanova EP, Rodriguez-Valera F. 2014. Genomes of *Alteromonas australica*, a world apart. *BMC Genomics* 15:483. <https://doi.org/10.1186/1471-2164-15-483>.
 66. Pernthaler A, Pernthaler J, Eilers H, Amann R. 2001. Growth patterns of two marine isolates: adaptations to substrate patchiness. *Appl Environ Microbiol* 67:4077–4083. <https://doi.org/10.1128/AEM.67.9.4077>.