Evolutionary Ecology of the Marine Roseobacter Clade

Haiwei Luo, Mary Ann Moran
Department of Marine Sciences, University of Georgia, Athens, Georgia, USA

SUMMARY
Members of the Roseobacter clade are equipped with a tremendous diversity of metabolic capabilities, which in part explains their success in so many different marine habitats. Ideas on how this diversity evolved and is maintained are reviewed, focusing on recent evolutionary studies exploring the timing and mechanisms of Roseobacter ecological diversification.

INTRODUCTION
The major clades of heterotrophic bacteria inhabiting surface ocean waters were initially discovered with culture-independent 16S rRNA gene surveys (1, 2) and found to include the alphaproteobacterial SAR11, Roseobacter, and SAR116 clades and the gammaproteobacterial SAR86 clade, among others (1–3). Studies over the next 2 decades revealed enormous genomic and physiological diversity among these major clades, correlating with differences in genome size, gene repertoire, G+C content, ecological strategy (free living, patch adapted, or eukaryote associated), and trophic strategy (heterotrophy, phototroph, or autotrophy). Such diversity is relevant from an ecological perspective because it determines the roles of these bacterial lineages in oceanic elemental cycles and their interactions with marine eukaryotes. However, how this diversity evolved and is maintained in a well-mixed seawater matrix has only begun to be addressed (4–8). In efforts to understand the ecology of marine bacteria and predict their responses in a changing ocean, the history and mechanisms of genome change are crucial puzzle pieces. In this review, we focus on the evolutionary processes underlying the ecological roles of the marine Roseobacter clade.

The Roseobacter clade is found predominantly in marine environments, representing up to 20% of bacterial cells in some coastal ecosystems and 3 to 5% of bacterial cells in open ocean surface waters (9). In addition to living freely in bulk seawater (10, 11), roseobacters are dominant members of the bacterial communities associated with phytoplankton (12–17), macroalgae (18, 19), and various marine animals (20–23). Both mutualistic (24, 25) and pathogenic (26–29) life-styles have been suggested. Clade members are ubiquitous in temperate and polar oceans (30); the latter include sea ice habitats, in which Roseobacter is a major bacterial phylotype (31, 32). Roseobacters are also abundant in coastal sediments (33), deep pelagic ocean (34), and deep-sea sediments (35).

Since the publication of the first Roseobacter genome from Ruegeria pomeroyi DSS-3 (formerly Silicibacter pomeroyi DSS-3) a decade ago (36), numerous other Roseobacter genome sequencing projects have been launched. To date, genomes of >50 isolates and 4 uncultivated single cells have been sequenced. An important lesson learned through genome analyses is that members of the Roseobacter clade are equipped with a tremendous diversity of metabolic capabilities and regulatory circuits (9, 25, 36–39), which in part explains their success in a variety of marine habitats. Ideas on how this diversity evolved and is maintained are reviewed, focusing on recent evolutionary studies yielding information on the timing and mechanisms of ecological diversification.

WHAT DEFINES A ROSEOBACTER?
Members of the Roseobacter clade form a subgroup of the Rhodobacterales that shares ≥89% identity of 16S rRNA gene sequences (40). Most of the cultured members have large genomes, high
G+C content (60% ± 4%), and versatile metabolic capabilities. An interesting exception is a cultured member branching at the base of the Roseobacter phylogeny, strain HTCC2255, which has a streamlined genome, low G+C content (37%), and a paucity of genes for transcriptional regulation, motility, and cell-cell interactions (41). A few uncultured Roseobacter single-cell amplified genomes (SAGs), although evolutionarily divergent from the basal HTCC2255 strain, also have streamlined genomes and low G+C content (39% ± 1%) (42).

Despite this broad genomic diversity, roseobacters comprise a well-supported clade in the Alphaproteobacteria tree. The strong evolutionary coherence of roseobacters originally discovered by 16S rRNA gene analysis (40) has been validated recently by applying phylogenomic approaches on concatenated protein data sets consisting of conserved single-copy orthologous sequences (37, 43). Concern over potential systematic errors in the data set (for example, due to the extreme differences in G+C content among Alphaproteobacteria taxa) motivated the use of multiple models and computational approaches, including a model correcting for amino acid compositional bias (44), a model integrating heterogeneity in the amino acid replacement process across sites of a protein alignment (45), a model accounting for variation of the substitution rate at a site across time (46), and a gene partition model considering heterogeneity among genes based on a gene partition framework (47). These approaches all give strong support for a monophyletic clade of roseobacters, as does an analysis of nearly 200 gene trees for orthologous gene families conserved in Alphaproteobacteria (H. Luo, P. G. Foster, and M. A. Moran, unpublished data). Therefore, phylogenetic coherence, and not necessarily conserved metabolic capabilities or phenotypic traits, defines a roseobacter.

While the Roseobacter clade itself is strongly supported, the within-group phylogeny based on the most commonly used approach of 16S rRNA gene analysis is not (40). One consequence of this is that the current Roseobacter taxonomy is at odds with the phylogeny (37, 41–43). For instance, the genera Ruegeria, Roseobacter, Phaeobacter, and Oceanicola are clearly paraphyletic lineages. This noncoherent assignment of scientific names can cause confusion when interpreting the ecology and evolutionary biology of the group.

ECOLOGICAL AND EVOLUTIONARY GENOMICS OF CULTIVATED ROSEOBACTERs

Genome Content

Roseobacter genome content has been addressed in several reviews (9, 37, 48, 49), and here we update the current status by including ~15 genome sequences that have recently become publically available. Based on 105 single-copy genes that are present in all the Roseobacter closed genome sequences (37), the currently sequenced Roseobacter isolates have at least 91% coverage and a median of 100% coverage (42).

An unusual genetic capability characteristic of roseobacters is the alphaproteobacterial gene transfer agent (GTA), first discovered in Rhodobacter capsulatus (RcGTA) (50). The RcGTA consists of 15 to 17 cotranscribed genes, some of which show homology with phage genes (51). While phage particles package complete phage genomes with occasional inclusion of host bacterial DNA, GTA particles randomly package bacterial DNA only (52). They have been shown to mediate high rates of gene transfer, as much as 10⁶-fold higher than estimated transformation and transduction rates (53). Among marine bacterial lineages, the Roseobacter clade is the only one to carry GTA genes, and our survey of 52 cultured Roseobacter genomes showed that 46 of them contain a complete or nearly complete set (Fig. 1). Three clade members have been experimentally demonstrated to produce RcGTA-like particles, including Ruegeria pomeroyi, Roseovarius nubinhibens, and Ruegeria mobiliis (53, 54). Diverse Roseobacter-like GTA genes can be amplified from estuarine waters (161), yet they are rarely observed in ocean metagenomes where amplification is not employed, including the Global Ocean Sampling (GOS) data set (54). Thus, the abundance and distribution of GTA genes in marine ecosystems remain unclear, as does their role in mediating lateral gene transfer (LGT) during Roseobacter evolution.

Members of the Roseobacter clade have a variety of mechanisms for obtaining energy. In addition to heterotrophic energy acquisition through the oxidation of organic matter, genome analysis indicates that some members are also capable of phototrophy. Light utilization involving bacteriochlorophyll a synthesized by aerobic anoxygenic phototrophic (AAP) roseobacters is found in 14 phylogenetically diverse strains (Fig. 1). Light utilization based on proteorhodopsin is found only in the streamlined strain HTCC2255 (Fig. 1), and that based on xanthorhodopsin is present in three strains (Fig. 1). For all phototrophic roseobacters, genes for carbon fixation are absent, and thus, these processes may produce energy but do not directly provide fixed carbon; there has been speculation, however, of heightened anaplerotic pathways fueled by light (55). Nonobligate chemolithotrophy is also a common signature in Roseobacter genomes. Twenty-four of the 52 strains have the form I version of carbon monoxide dehydrogenase (Fig. 1), and CO oxidation has been demonstrated experimentally for several strains (36, 56). Thirty-six genomes carry the Sox genes that mediate the oxidation of sulfide or thiosulfate (57) (Fig. 1) and may also provide a source of energy. Among the two types of dissimilatory nitrate reductases (periplasmic napA and cytoplasmic narG) and dissimilatory nitrite reductases (nirK and nirS), about half of the 52 strains contain at least one of these genes (Fig. 1). The taxonomic distribution of these chemolithotrophic gene sets does not map coherently to species phylogeny, suggesting that gain and loss of these traits have occurred multiple times during Roseobacter evolution, consistent with the scenario that roseobacters have continuously explored new ecological habitats. A few of these ecologically relevant genetic traits, including complete photosynthetic gene clusters (58), capabilities for production of the antibiotic tropodithietic acid (TDA) (59, 60), and type IV secretion systems (61), are found on plasmids, suggesting that extrachromosomal DNA could be an important mechanism by which genes useful for environmental adaptation are transferred among roseobacters.

Genome Evolution

By using a likelihood-based ancestral genome content reconstruction method (62), gene families summarizing all protein-coding genes in 39 Roseobacter genomes and 26 other genomes representing major lineages in the Alphaproteobacteria were mapped to their ancestral nodes in the Alphaproteobacteria tree (41). In general, early evolution in the Roseobacter clade was predicted to experience a net genome reduction from a large common ancestral genome, followed by two episodes of genome innovation and ex-
FIG 1 Survey of select genes and metabolic pathways in 52 *Roseobacter* isolate genomes. % complete, estimate of genome completeness; GTA, gene transfer agent; AAP, aerobic anoxygenic phototrophic; PR, proteorhodopsin; XR, xanthorhodopsin; *nasA*, assimilatory nitrate reductase; *nirB*, assimilatory nitrite reductase; *napA*, periplasmic dissimilatory nitrate reductase; *narG*, dissimilatory nitrate reductase; *nirS*, dissimilatory nitrite reductase; *dmdA*, dimethylsulfoniopropionate demethylase; *soxB*, sulfur oxidation gene; *coxL* type I, carbon monoxide oxidation; B7, biotin synthase; B1, thiamine synthase; B12, cobalamin synthase; Type IV Sec, type IV secretion system. Colors indicate four major clades of isolate genomes. The phylogenetic tree was constructed based on a concatenation of ~50 single-copy conserved protein sequences using the RAxML software.
pansion through lateral gene transfer (41) (Fig. 2). Although other authors analyzing single-gene trees suggested that roseobacters might have arisen more anciently (63), molecular dating analysis based on the currently available genomes and using a relaxed molecular clock model predicts that the Roseobacter ancestor existed around 260 million years ago (mya) (Fig. 2). A first predicted episode of genome expansion shortly after the emergence of this Roseobacter ancestor shows that it would have coincided with the rise of two red-plastid-lineage eukaryotic phytoplankton groups, the dinoflagellates and coccolithophorids, around 250 mya (41) (Fig. 2).

One explanation for the predicted coincidence of Roseobacter genome expansion and eukaryotic phytoplankton diversification is that the eukaryotes provided new ecological habitats for ancestral roseobacters (41). Indeed, modern lineages of roseobacters are abundant and consistent components of the phycosphere community of these phytoplankton groups (12, 15–17, 64–67). The predicted occurrence of genes involved in motility and chemotaxis in the ancestor of the Roseobacter clade shows that it would have coincided with the rise of two red-plastid-lineage eukaryotic phytoplankton groups, the dinoflagellates and coccolithophorids, around 250 mya (41) (Fig. 2).

A second episode of genome innovation during the evolution of the Roseobacter clade is predicted to have occurred more recently (Fig. 2), based on elevated LGT rates calculated for several leaf branches (41). The basal lineage consisting of reduced genomes related to Roseobacter member strain HTCC2255 appears to have escaped both of these expansion events and become streamlined directly from the common ancestor of the clade (41). Thus, not all extant roseobacters are predicted to be descendants of the ancestral lineages that experienced genome innovation.

The evolutionary mechanisms underlying the diversification of roseobacters can be compared to those driving the evolution of another dominant surface ocean lineage that shares a common ancestor in the Alphaproteobacteria, the SAR11 clade. In contrast to Roseobacter genomes, SAR11 genomes are uniformly small, have low G+C content (~30%), and harbor limited metabolic capabilities (72, 73). Ecologically, the roseobacters have been proposed to represent a patch-adapted ecological strategy that takes advantage of seawater microenvironments with elevated nutrient

![FIG 2 Evolutionary time line of the Roseobacter lineage. (Left) The first predicted episode of Roseobacter genome expansion coincided with the radiation of marine dinoflagellates and coccolithophorids ~250 mya. The diversification of diatoms occurred more recently. (Middle) Free-living and particle-associated roseobacters cooccur in the surface ocean, each with a distinct set of traits. (Right) In a future ocean, greater dominance by phytoplankton lineages with smaller cell sizes could lead to a decrease in particle-associated roseobacter populations.](http://mmbr.asm.org/)

concentrations, while members of the SAR11 clade are considered to use a free-living planktonic strategy adapted to nutrient-depleted bulk seawater (36, 72, 74). The SAR11 clade may also have a much longer evolutionary history, inhabiting the ocean for >800 million years (41). In contrast to roseobacters, there is currently no evidence for major genome innovations in SAR11, and genome size changes over the evolution of this group suggests that it evolved at a lower rate than did the roseobacters (41). The common ancestor of the SAR11 clade was predicted to have contained only ~2,000 genes (41). For the Roseobacter clade, gene gain appears biased toward transcriptional regulators, replication/recombination/repair genes, and defense mechanism genes; such traits might be important for roseobacters to compete well with cooccurring microbial populations on particles and living surfaces in the marine environment. For SAR11, gene gain is predicted to have been biased toward genes for cell wall biogenesis and pilus synthesis (41).

ECOLOGICAL ASSOCIATIONS BETWEEN ROSEOBACTERS AND EUKARYOTES

Roseobacter-Phytoplankton Interactions

Roseobacter-phytoplankton interactions are likely to be based on acquiring organic matter on the bacterial side and obtaining essential vitamins and regenerated inorganic nutrients on the phytoplankton side (24). Roseobacters may also function as probionts that deter algal pathogens (24). As these interactions are potentially relevant for explaining evolutionary relationships between roseobacters and phytoplankton, we summarize known interaction mechanisms and chemical exchanges. Both genomic and experimental evidence suggests that some interaction mechanisms are common to most members of the Roseobacter clade and most types of eukaryotic phytoplankton, while others are limited to specific phylotypes of each. General attributes that may be useful in interactions with eukaryotic organisms include motility and chemotaxis (49, 75), type IV secretion systems (9, 76), quorum-sensing systems (77), and probiotic biosynthesis (78).

Among the list of organic compounds released by phytoplankton that have been shown experimentally to be assimilated by roseobacters is the two-carbon compound glycolate, which is excreted during autotrophic photosynthesis (79, 80) and can comprise 10 to 50% of the phytoplankton exudate in marine environments (80–82). Roseobacters are among a select group of marine bacterial lineages capable of metabolizing this compound (83, 84). Carbohydrates are also exuded by phytoplankton, with the composition possibly varying among lineages. Roseobacter clade members not only grow on these carbohydrates (e.g., Planktotaenial frisia) but also may grow better on exudates produced by certain species (85), raising the possibility of an adaptive association of roseobacters with specific phytoplankton lineages. The organic sulfur compound dimethylsulfoniopropionate (DMSP) is produced in abundance by dinoflagellates and coccolithophorids (12, 17, 67), two phytoplankton groups often found associated with roseobacters in ocean waters (15, 16, 64–66). DMSP acts as a specific chemical cue that attracts motile and chemotactic bacteria, including roseobacters (86). The fact that the Roseobacter clade is only one of two marine bacterial lineages harboring both of the known pathways for DMSP degradation (the other being the SAR116 clade) (87) suggests that this compound may be a particularly important currency in Roseobacter-phytoplankton interactions (75). Finally, polyunsaturated aldehydes (PUAs) produced mainly by diatoms may selectively inhibit Roseobacter members, with evidence for poor growth in the presence of PUAs for strains related to Roseobacter litoralis and Phaeobacter gallaeciensis (88).

Interactions based on the release of metabolites are not a one-way street, as roseobacters may also release compounds that affect phytoplankton physiology. A recent study showed that exudates from the Roseobacter clade member Dinoroseobacter shibae stimulate the metabolism of the diatom Thalassiosira pseudonana, including enhanced production of picolinic acid, which can form metal complexes with limiting trace metals such as iron (89). The ability to synthesize the soluble vitamins B1 (thiamine), B2 (riboflavin), and B12 (cobalamin) is often lacking in phytoplankton, with ~50% of >300 eukaryotic phytoplankton species being auxotrophic for vitamin B12, ~25% being auxotrophic for vitamin B1, and 8% being auxotrophic for vitamin B2 (90). Every one of the 52 Roseobacter genomes has evidence of a functioning vitamin B12 biosynthesis pathway (90) (Fig. 1), even though >60% of marine bacterial species appear to be unable to make this vitamin. Over half of the sequenced Roseobacter strains also have genes for biosynthesis of vitamins B1 and B2 (Fig. 1). Whether metabolite exchanges between marine bacteria and phytoplankton rise to the level of symbiosis is a matter of debate (91, 92), but species-specific mutualism is not required for such interactions to have evolutionary consequences.

On the pathogenic side of Roseobacter-phytoplankton interactions, Phaeobacter gallaeciensis has been shown to produce an algalicid compound against senescence Emiliania huxleyi populations and may help to terminate blooms (27, 28). A similar pathogenic function was also found for a Roseobacter clade-affiliated (RCA) cluster strain, LE17, which demonstrated algalicidal activity against dinoflagellates (93). A set of genes potentially noteworthy for interactions with phytoplankton are those of the type IV secretion system, noticed in one of the first Roseobacter genome sequences (36) and now known to be present in nearly half of the strains surveyed (Fig. 1). The roles of these systems in roseobacters have yet to be deciphered, although studies of other model bacteria (e.g., Agrobacterium) showed that they mediate transfer of DNA and proteins to eukaryotic cells and thereby modulate physiological processes in the recipient organism (94).

Roseobacter-Coral Interactions

Roseobacters are also known to associate with a variety of marine macroalgae and animals, including green and red seaweed (18, 19), sponges (95), squid (22), fish larvae (96), sea urchins (21), and oysters (26, 97). A few of them have been demonstrated to induce physiological changes in the host, both positive and negative, including gall disease in the red alga Prionitis lanceolata (98), juvenile oyster disease (99), infection of the red alga Delisea pulchra (100), and inhibition of the settlement of fouling organisms on the green alga Ulva australis (19). The most extensive research, however, has focused on associations with corals (101–107), as roseobacters can account for up to 50% of the coral mucus microbiome (108–110). It has been hypothesized that DMSP and dimethylsulfide (DMS) are the two compounds that structure the community of coral-associated microbes, since isolates from these populations are enriched with the ability to degrade DMS/DMS (111). As is the case for Roseobacter interactions with other eukaryotes, it is not yet clear whether a physical association signifies a specific mutualistic interaction. There is evidence that roseobacer-
ters may play a role in coral reproduction, based on both consistent associations between the bacteria and early developmental stages of corals (112, 113) as well as increases in Roseobacter abundance after coral mass spawning (114). Roseobacters may also have a role in coral bleaching by acting as probiotics against pathogens (115, 116). Some Roseobacter lineages have been found to be significantly associated with lesioned corals (106, 117, 118). Whether interactions with corals are specific or general and whether the outcomes are beneficial or pathogenic, the associations are nonetheless likely to have shaped aspects of Roseobacter evolution.

ECOLOGICAL AND EVOLUTIONARY GENOMICS OF UNCULTIVATED ROSEOBACTERS

Genome Analyses

Many major Roseobacter lineages, including those that are most numerically dominant in the ocean, do not yet have cultivated representatives. Metagenomic data sets are valuable resources for studying these uncultivated roseobacters, as long as the taxonomic classification of metagenomic reads is unequivocal. Unfortunately, the widely applied approach to metagenomic analysis of using BLAST best hits can produce an unacceptable level of false-positive results due both to uncorrected genetic distances of similarity scores and the biased taxonomic composition of the database (119). De novo genome assembly from deeply sequenced metagenomes is one way around this problem, and indeed, a near-complete Roseobacter genome closely related to strain HTCC2255 was assembled from a sample from coastal seawater near Seattle, WA (120). When roseobacters are not dominant or when the sequencing depth of a metagenome is shallow, however, an alternate approach uses the computed patterns in \( d_N \) the number of non-synonymous (amino-acid-changing) substitutions per nonsynonymous site. A two-step bioinformatics procedure first assigns metagenomic gene fragments to a clade based on the values of \( d_N \) computed between the metagenomic gene fragment and its orthologs in cultured members and then uses the paired-end reads of these conservatively assigned metagenomic sequences as a homology-independent survey of biological function (43). Applied to the ~7 million Sanger reads in the 2007 GOS metagenomic data set, the method identified ~3,000 unambiguous Roseobacter reads exhibiting several systematic differences in genome content compared to their cultured counterparts. The metagenomic Roseobacter data set had fewer genes for signal transduction and cell surface modifications but more genes for Sec-like protein secretion systems (43). Interestingly, these gene functions have been suggested to be signatures for distinguishing patch-adapted from free-living ecological strategies in marine bacteria (121, 122). One conclusion arising from the analysis of metagenomic data is that genomic analyses focusing on cultured roseobacters is biasing our view of the lineage’s ecology (43).

Another approach to understanding Roseobacter biology without the bias of cultivation is the use of single-cell genome sequencing. Recently, four SAGs of uncultivated roseobacters were obtained from surface waters of the North Pacific, South Atlantic, and Gulf of Maine (123). Phylogenomic analyses showed that 3 of the SAGs comprise a novel clade (the SAG-O19 clade) in which no cultured representatives have been found. Moreover, this clade appears to represent up to 35% of the Roseobacter sequences in samples from surface ocean waters in the GOS (42). These SAGs have low G+C content (39% ± 1%) and a reduced percentage of noncoding DNA (7.5% ± 0.4%) and are predicted to have streamlined genomes (2.6 to 3.5 Mbp). These characteristics are consistent with those of the oceanic roseobacters identified by the \( dN \) pipeline, and they show the same systematic divergences from cultured strains (42) (Fig. 3).

Evolution of an Enigmatic Roseobacter Lineage

As the SAG-O19 clade represents an abundant extant lineage with low G+C content, it may be informative to consider how the evolutionary forces affecting this clade compare to mechanisms invoked for other G+C-poor bacterial genomes. In a situation where synonymous (silent) nucleotide sites are saturated with substitutions, as is the case for the three SAG-O19 clade members, an alternate population genetic approach based on the ratio of radical to conservative nonsynonymous nucleotide substitution rates (\( d_{dR}/d_{dC} \) ratio) can be used instead, based on classification of amino acids according to physicochemical properties such as charge (positive/negative/neutral) (124, 125). By using this approach, a higher mean \( d_{dR}/d_{dC} \) ratio was found for >500 orthologous genes in the uncultivated SAG-O19 clade than in other Roseobacter lineages. Theory predicts that such inflated \( d_{dR}/d_{dC} \) ratios occur when purifying selection is not efficient and thus argues that genetic drift may have played a prominent role during the evolution of the SAG-O19 clade (42). Such an interpretation might be...
problematic if the selective pressure to decrease G+C content and, hence, nitrogen (N) usage in the proteome (126) interferes with the selective pressure for conservation of amino acid physicochemical properties. In other words, the fitness cost of radical changes could be balanced by the benefit of using less N in both genomes and proteomes (127). However, two observations that cast doubt on the idea that low G+C content is under particularly strong selection in marine bacteria are that freshwater relatives of the low-G+C-content SAR11 clade similarly have low-G+C-content genomes, even though N is not limiting in freshwater environments (128), and that both low-G+C- and high-G+C-content lineages coexist in many major marine clades, including not only the roseobacters but also the SAR116 lineage and several marine Gammaproteobacteria groups.

Based on recent studies showing that mutational bias from G/C to A/T is universal among bacterial genomes (129–132), one explanation for the observed paucity in G+C-rich codons in genomes of the SAG-O19 clade is that an inefficiency of purifying selection under such a mutational bias would accelerate the replacement of G+C-rich codons. Thus, although counter to the more established theory that selection has been the exclusive force driving the evolution of surface ocean bacterial populations (8, 133), there are also reasonable arguments in favor of the inefficiency of purifying selection playing a role, at least in the evolution of SAG-O19 genome composition. Regardless of the mechanism, it is clear that different evolutionary processes have been at work in the SAG-O19 clade compared to its cultured Roseobacter relatives.

Inclusion of Uncultured Roseobacters in Phylogenomic Analyses

With the availability of genomic sequences of single cells, uncultivated bacterial and archaeal lineages have been regularly placed in phylogenetic trees along with their cultured relatives (134, 135). However, when the three uncultivated genomes in the SAG-O19 clade (SCGC AAA015-O19, SCGC AAA298-K06, and SCGC AAA300-T04) are included in phylogenetic trees based on standard approaches, such as multiprotein concatenation using maximum likelihood (134, 135), the branching order of several major clades is no longer supported (42). The substantially different G+C contents of the 3 SAGs compared to those of most isolates (39% ± 1% versus 60% ± 4%) may be one cause of this problem, a phenomenon known as nonstationarity (136–140). Indeed, posterior predictive simulation assessing the fit of homogeneous versus heterogeneous models to each data set (i.e., models that do not account for compositional heterogeneity in G+C content versus those that do) showed that half of the orthologs used for tree building significantly violated the composition-homogeneous assumption. Since many standard phylogenetic programs cannot account for compositional heterogeneity, the node-discrete composition heterogeneity (NDCH) and node-discrete rate matrix heterogeneity (NDRH) models that allow for variation in G+C contents and rate matrices across branches were used instead (44). These models resolved the 3 SAGs into a well-supported clade while maintaining the evolutionary relationships of cultured lineages identified previously (42). An understanding of the evolutionary history of the marine Roseobacter clade requires the inclusion of the uncultured members, even if they pose computational challenges, particularly given the genomic evidence that they represent very different organisms and life history strategies. It can be envisioned that this nonstationary phylogenetic model will become increasingly useful as additional SAGs fill in our understanding of Roseobacter diversity and ecology.

Genome Streamlining in the Roseobacter Clade

Genome streamlining is a well-known process in obligate intracellular bacteria. For these taxa, the primary mechanism of gene loss is thought to be relaxation of purifying selection on genes for which host proteins perform the same biological function, rendering mutations selectively neutral (141). Additionally, obligate intracellular bacteria are assumed to have small effective population sizes as a result of frequent population bottlenecks, leading to fixation of slightly deleterious mutations through genetic drift (142, 143). Genome streamlining is also recognized as a prevalent feature among diverse lineages of free-living planktonic marine bacteria (123). A few prominent examples are the alphaproteobacterial SAR11 clade (72), the gammaproteobacterial SAR86 clade (144), and the cyanobacterial Prochlorococcus clade (145, 146). The primary evolutionary mechanisms of streamlining proposed for marine bacteria is selection for metabolic efficiency (72, 147). Secondarily, these bacteria also typically have reduced cell sizes and thus can maximize surface-to-volume ratios to facilitate nutrient acquisition (72, 121). However, the current theory cannot explain why bacteria with average-sized genomes cooccur with bacteria with streamlined genomes in oligotrophic seawater and even represent dominant groups in some oceanic habitats. This “paradox of genome streamlining” has been perplexing (73). Although the marine Roseobacter clade was initially thought to contain mainly members with large genomes, both metagenomic approaches and single-cell genome sequencing indicate that for many oceanic roseobacters, this is not the case (42). Genome streamlining is predicted to have occurred during the evolution of the basal lineage of the clade, represented by isolate HTCC2255, with a nearly complete genome of 2.54 Mb, and a highly related single-cell genome, with ~80% coverage and an estimated genome size of 2.64 Mb. Ancestral genome content reconstruction and COG functional category analysis showed that HTCC2255 shares some similarities in genetic architecture with the common ancestor of the SAR11 clade, both of which are more similar in gene content to the extant SAR11 genomes than to extant Roseobacter genomes (41). This convergent evolution of genetic makeup suggests that the basal HTCC2255 clade is consistent with a free-living planktonic strategy. In the predicted streamlining from the Roseobacter common ancestor that led directly to the HTCC2255 clade, genes that were lost were biased toward functions involving cell–cell interactions, transcriptional regulation, and motility (41). While these biological functions are likely essential for cells living on surfaces or switching between free-living and surface-associated strategies, they may become dispensable and expensive to maintain when cells grow singly in bulk seawater. One explanation for functionally biased gene loss could be inefficiency of purifying selection due to restriction of habitats (population bottleneck) or exploration of new habitats (founder effect), resulting in the preferential loss of nonessential genes. While comparisons of ancestor-descendant genome contents have yielded useful information regarding the evolutionary mechanism giving rise to the basal HTCC2255 clade, a fuller resolution of the evolutionary path of this basal lineage during the past 260 million years is lacking.

In the case of the SAG-O19 clade, all three genome sequences...
were only partially recovered during single-cell sequencing, yet several independent lines of evidence support genome streamlining. First of all, genome size estimates based on single-copy genes suggest that SAG-O19 clade members are smaller than most cultured roseobacters (see above). Second, nonobligate endosymbiotic bacterial genome reduction often occurs with a reduction in the percentage of noncoding DNA (72), a feature which is also observed for these SAGs (Fig. 3). Finally, gene content characteristics attributed to streamlined marine bacterial genomes are found in the SAG-O19 clade, including an underrepresentation of genes involved in transcriptional regulation and replication/recombination/repair (41, 121, 122) relative to roseobacters with larger genomes. Since a substantial part of the SAG genomes was not sequenced (45 to 77%), it is more challenging to reconstruct the evolutionary pathway giving rise to this clade. If inefficiency of purifying selection drove the evolution of the HTCC2255 clade, as suggested above, it is plausible that it also played an important role in genome streamlining of this clade. Interestingly, the *Roseobacter* R11 lineage that has been observed to live endosymbiotically with the marine macroalga *Delisea pulchra* has not undergone genome reduction, having a genome size of 3.93 Mb (42). This is expected, however, since as the bacterium is facultative symbiont, it must maintain a genome repertoire sufficient for independent living.

**A New Look at Population Size**

There are two ways of conceptualizing the population size of organisms: census population size ($N_c$), which is a headcount of a population and has significance from the ecological standpoint, and effective population size ($N_e$), which determines the capability for environmental adaptation and has relevance from an evolutionary viewpoint. The widely used GOS metagenomic data from ocean surface waters, which provide taxonomic information for environmental adaptation and has relevance from an evolutionary viewpoint. The widely used GOS metagenomic data from ocean surface waters, which provide taxonomic information free of the biases inherent in PCR-amplified 16S rRNA gene surveys, suggest a lower $N_c$ for roseobacters (4% of surface ocean bacteria) than for the SAR11 clade (31%) and several other bacterial taxa (SAR86, 6%; *Bacteroidetes*, 6%; *Actinobacteria*, 8%) (148).

However, the genomes sequenced in the GOS were primarily from small bacteria (i.e., those falling in the size range of 0.2 to 0.8 μm in diameter) at the expense of larger and particle- or host-associated cells, instituting a bias that may affect estimates of census population size. To better address the $N_e$ of the *Roseobacter* clade, we analyzed new marine metagenomic data from the Southern Ocean (149) and Monterey Bay (A. Z. Worden, unpublished data), in which both free-living (0.1- to 0.8-μm) and particle-associated (0.8- to 3.0-μm) communities were sequenced. Based on queries of >2,000 orthologous gene families, the metagenomic data sets showed that roseobacters accounted for a higher percentage of the genomes in particle-associated communities than in free-living communities by >2-fold overall (Fig. 4), with 11 of 13 Southern Ocean samples and 3 of 4 Monterey Bay samples following this pattern ($P < 0.001$ by $χ^2$ test). Thus, calculation of the census population size of the *Roseobacter* clade may be a sizeable underestimate in cases where sampling protocols systematically exclude larger and particle-associated cells.

From an evolutionary perspective, it is the $N_e$ that determines whether a deleterious mutation can be effectively purged from the population and whether a favorable mutation can be effectively spread, two major ways in which populations adapt to a changing environment (150). In a population with a low $N_e$ value, there is a greater chance that deleterious mutations drift to fixation while favorable mutations drift to loss, both of which reduce fitness. Calculation of the absolute values of $N_e$ requires knowledge of the mutation rate for a given population, but this has not yet been measured for any marine bacterial population (151). Streamlined marine bacteria that grow slowly and do not form colonies on solid media pose a particular challenge for approaches that rely on direct measurements of the mutation rate for the calculation of $N_e$ (151).

Alternatively, population genetic theory provides a method for comparing the relative values of $N_e$ based on the expectation that a population with a reduced $N_e$ will have an increasing number of slightly deleterious mutations that drift to fixation (152). Since nonsynonymous mutations are more likely to be slightly deleterious, it is expected that the ratio (ω) of the number of nonsynonymous substitutions per nonsynonymous site ($d_{ns}$) to the number of synonymous substitutions per synonymous site ($d_s$) is greater in populations with a lower $N_e$ value (153). Applying this approach to surface ocean bacteria, comparisons of ω were made for >400 orthologous genes among cooccurring *Alphaproteobacteria* populations. *Roseobacter* lineages exhibited a range of relative $N_e$ values, including some populations whose substitution patterns produced higher calculated $N_e$ values than those for SAR11 populations (151). The apparent mismatch between $N_e$ and $N_c$ is accounted for to some extent by the undersampling of lineages that have larger cells and/or are associated with particulate material (Fig. 4). Other possible factors include variability in the fraction of cells that are inactive and the population history of the lineages (e.g., population bottlenecks) (154).

Overall, neither census population size nor life history strategy is a good predictor of effective population sizes, and direct measures of $N_e$ are ultimately necessary to understand the ability of *Roseobacter* populations to adapt to a changing ocean. Under the neutral theory model, the nucleotide diversity at synonymous sites in a bacterial population equals twice the product of its effective population size and spontaneous mutation rate (per generation). Thus, one way to directly estimate the absolute value of $N_e$ is to determine the mutation rate through a mutation accumulation experiment in which multiple parallel cell lines are bottlenecked to a single cell regularly over several thousand generations, as has been done for a few model bacteria (155) but not for any that are marine. *Roseobacter* lineages can serve as model organisms for this experiment, since many grow readily on solid media and are amenable to experimental evolution studies.

**CONCLUDING REMARKS**

Although the *Roseobacter* clade is among the most frequently sampled of surface ocean bacterial lineages, it is perhaps one of the least known due to the enormous diversity of clade members with regard to habitats and metabolic capabilities. Over 2 decades of research on this lineage have established its importance in global carbon and sulfur cycles. Many of the attributed biogeochemical functions appear to be linked to the ability of group members to physically associate with phytoplankton, corals, and other eukaryotes and affect eukaryote function as mutualists, probionts, and pathogens. How climate change may diminish or strengthen the abundance and activity of *Roseobacter* lineages is currently unknown.

Following the emergence of the *Roseobacter* ancestor, estimated to have appeared about 260 mya based on analysis of cur-
FIG 4 Roseobacter sequences in metagenomic data sets, calculated as a percentage of free-living (0.1 to 0.8 μm) (green bars) and particle-associated (0.8 to 3.0 μm) (blue bars) reads in DNA from the Southern Ocean (13 stations [Stn.]) and Monterey Bay, CA (4 stations). Asterisks indicate significant differences between size fractions (P < 0.05). Data were obtained by analyzing 2,235 orthologous gene families shared by at least 20 of the 40 Roseobacter isolate genomes.
rently available genomes, some lineages are surmised to have experienced adaptive genome expansion through LGT. The radiation is predicted to have coincided in time with the emergence of dinoflagellates and coccolithophorids, and these phytoplankton groups may have provided new habitats and new roles in oceanic elemental cycles. In addition to eukaryotic phytoplankton, a number of other marine eukaryotic organisms are associated with bacterial communities in which roseobacters dominate, although it remains unknown whether, when, and how these eukaryotes may have imparted on the evolution of roseobacters. Addressing this question will require the availability of more Roseobacter lineage sequences associated with specific eukaryotes and an improved understanding of the biology and ecology underlying Roseobacter-eukaryote interactions.

A few Roseobacter lineages adopted a free-living planktonic strategy or are evolving toward one. Representing the most abundant Roseobacter lineages among the free-living ocean bacteria, these groups appear to have experienced an evolutionary stage in which effective population sizes were reduced compared to those of other Roseobacter lineages. Their evolutionary modifications, such as reductions in genome size and G+C content, may therefore have occurred under relaxed selective constraints. Because of their distinct gene repertoire, these planktonic roseobacters are postulated to play roles in marine elemental transformations that differ from those of their eukaryote-associated relatives.

Molecular evolutionary approaches promise to generate insights into the history and mechanisms underlying the ecological divergence of dominant surface ocean bacterioplankton (41–43, 151, 156–159), with the caveat that different evolutionary approaches will have different time scales of resolution. We are just starting to appreciate the evolutionary patterns of the Roseobacter clade as a whole, and we still have limited knowledge of how the tremendous ecological diversity of individual members evolved. Previous genomic sequencing projects have focused on sampling taxa that diverged anciently, enabling the reconstruction of a nearly complete evolutionary history of roseobacters. Future studies will also obtain genomes of closely related lineages within species, which are required to understand the ecological and evolutionary factors driving ongoing speciation processes. In fact, investigation of strain-level genomic variation has already started for marine Vibrio lineages associated with various microscale ecological niches, and significant insights regarding how microbes diversify in the seemingly well-mixed ocean have been obtained (5, 7, 160). Two current constraints on the evolutionary analysis of the Roseobacter lineage are the limited number of genome sequences available for each specific ecological niche and the biases associated with focusing on readily cultured representatives. It is anticipated that Roseobacter isolates and single cells isolated from samples from various ocean habitats will continue to be sequenced and, along with metagenomic data sets, will fill crucial gaps in our understanding of the evolutionary ecology of this important marine bacterioplankton clade.

ACKNOWLEDGMENTS
We thank A. Z. Worden for providing the unpublished Monterey Bay metagenomic data set and C. English for assistance with graphics. This research was funded in part by the Gordon and Betty Moore Foundation through grant GBMF338.01 and the National Science Foundation through grant OCE-1259562.

REFERENCES


2011. Correlating carbon monoxide oxidation with cox

Downloaded from http://mmbr.asm.org on February 21, 2021 by guest


73. Polz MF, Hunt DE, Preheim SP, Weinreich DM. 2006. Patterns and mechanisms of genetic and phenotypic differentiation in marine mi-


Mary Ann Moran is a Distinguished Research Professor in the Department of Marine Sciences at the University of Georgia. She has earned degrees at Colgate University (B.A., 1977), Cornell University (M.S., 1982), and the University of Georgia (Ph.D., 1987). Dr. Moran is a fellow of the American Association for the Advancement of Science and the American Academy of Microbiology and a recipient of the American Society for Microbiology's D. C. White Research and Mentoring Award. She serves on the editorial boards of Applied and Environmental Microbiology, Environmental Microbiology, and mBio. Dr. Moran’s research program focuses on the genetic basis of bacterial sulfur and carbon cycling in the ocean, with the goal of understanding the role of marine bacteria in the productivity of the ocean and the formation and flux of climatically active gases. Her research uses molecular microbial ecology and ecological genomics approaches to explore bacterial processes and their regulation in seawater.

Haiwei Luo received his M.S. in marine microbial ecology and Ph.D. in molecular evolution and bioinformatics from the University of South Carolina. During his doctoral study, he employed genome rearrangement and molecular sequence-based approaches to resolve evolutionary relationships of microbial and animal species and used population genetic models to address how genetic diversity evolved. He is currently a postdoctoral researcher in the Department of Marine Sciences at the University of Georgia. His current research interest includes understanding the role of selection, drift, mutation, and recombination in giving rise to the genetic diversity of marine microbes and resolving the controversial evolutionary relationships of marine microbial lineages. He will begin a position in the School of Life Sciences at the Chinese University of Hong Kong in January 2015, where he will continue to employ phylogenomic and population genetic approaches to understand microbial evolution and adaptation in the ocean.