



Coastal Transient Niches Shape the Microdiversity Pattern of a Bacterioplankton Population with Reduced Genomes

Xiao Chu,^{a,b}  Xiaojun Wang,^{a,c} Lok Shan Cheung,^d Xiaoyuan Feng,^{a,c} Put Ang, Jr.,^e Shing Yip Lee,^a Sean A. Crowe,^{d,f,g}  Haiwei Luo^{a,b,c}

^aSimon F. S. Li Marine Science Laboratory, School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

^bState Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

^cShenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China

^dDepartment of Earth Sciences, The University of Hong Kong, Pok Fu Lam, Hong Kong SAR

^eInstitute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

^fThe Swire Institute of Marine Science, The University of Hong Kong, Cape d'Aguiar Road, Shek O, Hong Kong SAR

^gDepartments of Microbiology and Immunology, and Earth, Ocean, and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia, Canada

Xiao Chu and Xiaojun Wang contributed equally to this work. Author order was determined alphabetically.

ABSTRACT Globally dominant marine bacterioplankton lineages are often limited in metabolic versatility, owing to their extensive genome reductions, and thus cannot take advantage of transient nutrient patches. It is therefore perplexing how the nutrient-poor bulk seawater sustains the pelagic streamlined lineages, each containing numerous populations. Here, we sequenced the genomes of 33 isolates of the recently discovered CHUG lineage (~2.6 Mbp), which have some of the smallest genomes in the globally abundant *Roseobacter* group (commonly over 4 Mbp). These genome-reduced bacteria were isolated from a transient habitat: seawater surrounding the brown alga, *Sargassum hemiphyllum*. Population genomic analyses showed that: (i) these isolates, despite sharing identical 16S rRNA genes, were differentiated into several genetically isolated populations through successive speciation events; (ii) only the first speciation event led to the genetic separation of both core and accessory genomes; and (iii) populations resulting from this event are differentiated at many loci involved in carbon utilization and oxygen respiration, corroborated by BiOLOG phenotype microarray assays and oxygen uptake kinetics experiments, respectively. These differentiated traits match well with the dynamic nature of the macroalgal seawater, in which the quantity and quality of carbon sources and the concentration of oxygen likely vary spatially and temporally, though other habitats, like fresh organic aggregates, cannot be ruled out. Our study implies that transient habitats in the overall nutrient-poor ocean can shape the microdiversity and population structure of genome-reduced bacterioplankton lineages.

IMPORTANCE Prokaryotic species, defined with operational thresholds, such as 95% of the whole-genome average nucleotide identity (ANI) or 98.7% similarity of the 16S rRNA gene sequences, commonly contain extensive fine-grained diversity in both the core genome and the accessory genome. However, the ways in which this genomic microdiversity and its associated phenotypic microdiversity are organized and structured is poorly understood, which disconnects microbial diversity and ecosystem functioning. Population genomic approaches that allow this question to be addressed are commonly applied to cultured species because linkages between different loci are necessary but are missing from metagenome-assembled genomes. In the past, these approaches were only applied to easily cultivable bacteria and archaea, which, nevertheless, are often not representative of natural communities. Here, we focus on the recently discovered cluster, CHUG, which are representative in marine bacterioplankton communities and possess some of the smallest genomes in the globally dominant marine *Roseobacter* group.

Editor Stephen J. Giovannoni, Oregon State University

Copyright © 2022 Chu 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 Haiwei Luo, hluo2006@gmail.com.

The authors declare no conflict of interest.

Received 28 February 2022

Accepted 1 July 2022

Published 26 July 2022

Despite being over 95% ANI and identical in the 16S rRNA gene, the 33 CHUG genomes we analyzed have undergone multiple speciation events, with the first split event predominantly structuring the genomic diversity. The observed pattern of genomic microdiversity correlates with CHUG members' differential utilization of carbon sources and differential ability to explore low-oxygen niches. The available data are consistent with the idea that brown algae may be home to CHUG, though other habitats, such as fresh organic aggregates, are also possible.

KEYWORDS Roseobacter, CHUG, population genomics, microdiversity, streamlined genomes, *Sargassum*

Most microbial species show fine-scale genetic diversity, known as microdiversity. Increasing evidence supports the ideas that microdiversity shapes functional traits of microbial taxa (1), promotes the temporal and spatial prevalence of microbial taxa (2, 3), stabilizes microbial communities under changing environmental conditions (4, 5), and builds a connection between microbial diversity and ecosystem functioning (6). The 16S rRNA gene amplicon sequences are often used to characterize the microdiversity of microbial populations, and those varying at the single nucleotide level (known as amplicon sequence variants [ASVs]) are shown to be ecologically meaningful (3, 7). However, an appreciable amount of genomic microdiversity has been commonly found in microbial populations whose members share identical full-length 16S rRNA genes, some of which have diversified into genetically discrete and ecologically distinct subpopulations (8, 9). Hence, analysis of 16S rRNA gene amplicon sequences may miss fine-scale microdiversity and obscure the interpretation of ecologically relevant units. Moreover, 16S rRNA genes, or any other marker genes alone, cannot provide functional information, making them of limited utility in linking microdiversity to functional traits.

The culturing of many closely related microbes, followed by genome sequencing, is a standard approach by which to study fine-scale genomic microdiversity and to pair it with phenotypic variation. Further, the genome sequences of cultured members reveal evolutionary mechanisms that structure the microdiversity. For instance, the extent of microdiversity is impacted by how niche-specifying alleles spread through microbial populations (10), which depends on the balance between recombination and selection (11). A high genome-wide recombination rate together with small selective strengths of genetic variants may induce gene-specific selective sweeps, where niche-specifying alleles spread within a population by recombination, thereby leaving the diversity of the remaining genomic regions largely unaffected (11). In contrast, a low genome-wide recombination rate along with strong selective strengths may lead to genome-wide selective sweeps, where niche-specifying alleles spread over the population by clonal expansion, thereby purging the genetic diversity throughout the whole-genome (11).

For bacteria inhabiting marine environments, the use of this population genomics approach is in its infancy. It has been largely applied to lineages with large and variable genomes, such as the *Vibrio* (9, 12, 13) and *Ruegeria* lineages (14–16), that readily grow on solid media. In sunlit pelagic oceans, however, a typical bacterioplankton cell often carries a reduced genome (with an average genome size of 2.2 to 2.6 Mbp) (17). Because of the technical difficulty and complexity in culturing genome-reduced bacterioplankton members, population genomic analyses are rare. A few studies have instead turned to single-cell amplified genomes (SAGs) (18). Despite the appreciable insights achieved, SAGs are limited in population genomic applications, owing to their low completeness and high error rates (19), which disable the inference of gene gain and loss events and hinder the interpretation of nucleotide substitution patterns, respectively, when nearly identical genomic sequences are compared (20). Further, SAGs do not allow for the pairing of physiological assays and genetic manipulations with population genomic analyses.

Here, we isolated 33 strains affiliated with the recently discovered CHUG cluster, a

genome-reduced lineage (~2.6 Mbp) (21) of the globally abundant marine Roseobacter group, which typically contains copiotrophic members with large and variable genomes (>4 Mbp on average) that are commonly associated with phytoplankton groups in the pelagic ocean (22, 23). While CHUG genomes are larger than those of the most abundant marine bacterioplankton lineages, such as the alphaproteobacterial SAR11 clade (1.3 to 1.4 Mbp) and the cyanobacterial *Prochlorococcus* (1.6 to 1.8 Mbp), they are, nevertheless, among the smallest of the Roseobacter genomes. The CHUG members differ by up to 1.8% in their 16S rRNA gene sequences, and they together differ from their closest sister group (3.9 Mbp on average) by 3.5% in the same gene (21). An important consequence of genome reduction is that CHUG members have lost the ability of de novo synthesis of vitamin B₁₂ (21). This is unusual because the potential for vitamin B₁₂ synthesis is widespread among Roseobacter members (21, 22) and sets the ground for mutualistic interactions between Roseobacter members and phytoplankton groups that are commonly auxotrophic for vitamin B₁₂ (24–26), suggesting that phytoplankton may not be an important pelagic niche for CHUG (21). In the present study, we chose the ambient seawater of the brown algae *Sargassum hemiphyllum* as our study site because two of the eight published CHUG genomes (21) were isolated from the same site and because they show evidence of metabolic interaction with brown algae by utilizing L-fucose, the dominant structural monosaccharide composing the polysaccharide fucoidan present in the cell walls of brown algae, as their sole carbon source. The 33 CHUG isolates share (nearly) identical 16S rRNA genes, thereby providing a unique opportunity to explore mechanisms that shape the microdiversity of an important, genome-reduced, marine bacterioplankton lineage.

RESULTS AND DISCUSSION

Phylogenetic and genetic structures indicate sequential speciation events. The 33 CHUG isolates (Data Set S1a) share identical 16S rRNA gene sequences, except that HKCCA1312 and HKCCA1076 show 1 to 2 nucleotide differences from the remaining isolates, and all of them show a whole-genome average nucleotide identity (ANI) above 95.8% (Fig. S1). Both statistics are well beyond the thresholds widely used to define bacterial species: 98.7% (alternatively, 99.0% and 99.5%) at the 16S rRNA gene level (30) and 95% at the ANI level (31), suggesting that the 33 isolates fall into a single “species”, according to the operational definition of the term.

The clonal evolutionary history, reconstructed from genomic regions free from recombination, showed that the 33 isolates have diverged into multiple phylogenetic clusters (Fig. 1A). PopCOGenT divided the 33 isolates into five genetically isolated populations (M1, M2, M3, M4, and M5) (Fig. 1A) and further divided M5 into two subpopulations (M5S1 and M5S2), between which genetic isolation has not been completed. Integrating the phylogenetic and genetic structures of these CHUG members, we inferred that the earliest speciation event occurred between M1M2 and M3M4M5, and follow-on events include the M3M4-M5 split and the M1-M2 split among others. Previous studies were focused on how a single speciation event shapes the pattern of genomic microdiversity (8, 9, 32, 33), yet the effect of multiple successive speciation events on genomic microdiversity has rarely been investigated. Therefore, the population-level data set of CHUG not only allows learning some new strategies that a representative, genome-reduced Roseobacter lineage uses for pelagic niche adaptation but also provides an opportunity to link complex, extant genomic microdiversity to a series of past speciation events and disentangle how each speciation event contributed to the collective genomic microdiversity. It is worth noting that “speciation” refers to a process of genetic separation, and this term is commonly used in studies of microbial population differentiation within an operationally defined species (32, 34, 35).

Genetic variations can be sourced from recombination and mutation, whose relative frequencies (ρ/θ) were estimated by ClonalFrameML (36). The entire CHUG population, comprised of 33 isolates, features a low recombination rate ($\rho/\theta = 0.079$), indicating either a reduced gene flow between the genetically isolated populations defined by

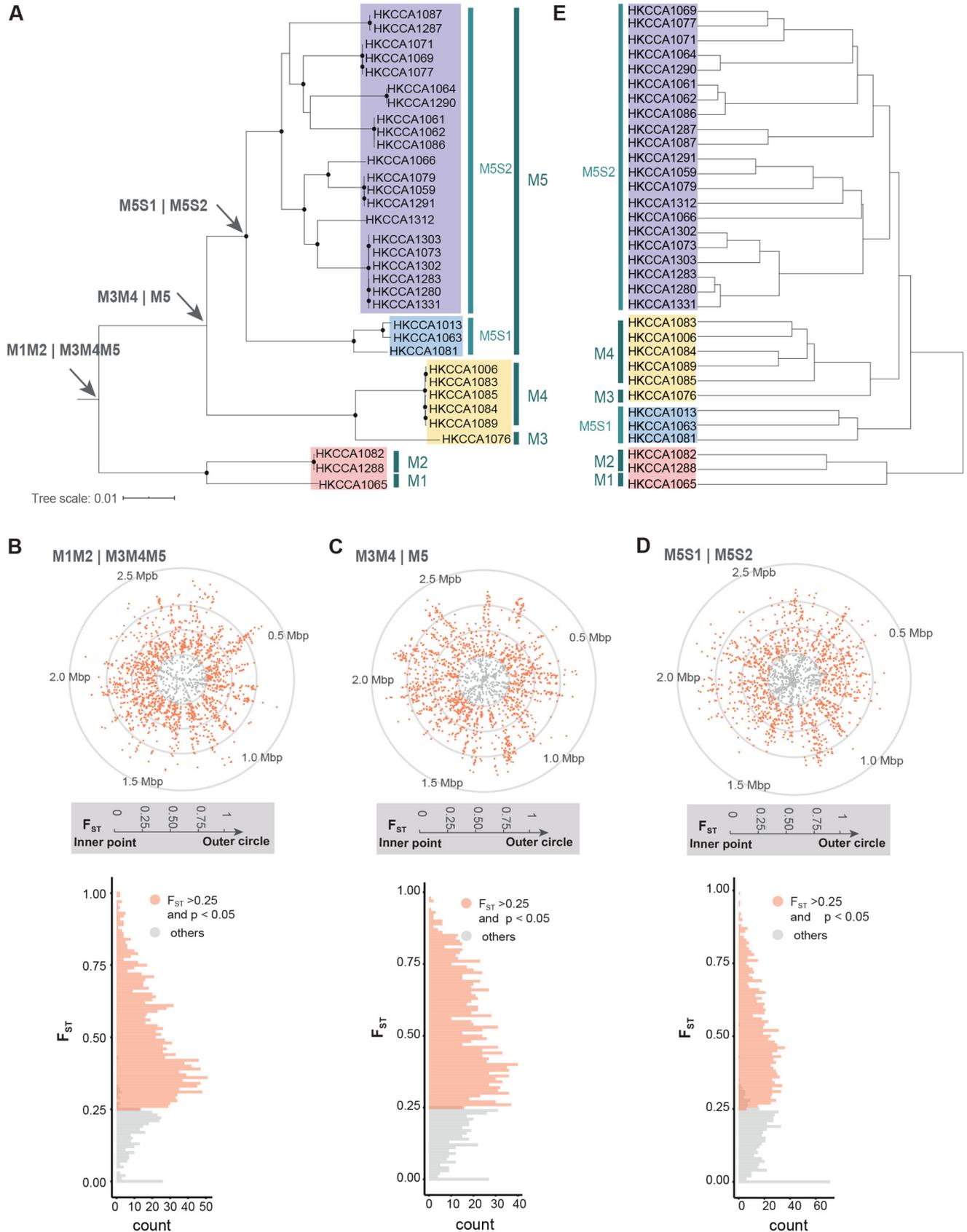


FIG 1 Population differentiation of the 33 CHUG isolates from a single 50 mL seawater sample surrounding a brown alga. (A) Genetically isolated
(Continued on next page)

PopCOGenT or that these members are inherently clonal (i.e., recombination rarely occurs). To test these competing hypotheses, we turned to the subpopulation M5S2, which is comprised of 21 members, displays a large amount of genetic diversity, and is panmictic by definition. The ρ/θ ratio of M5S2 (0.067) is comparable to that of the entire CHUG population, suggesting that clonality, rather than population differentiation, is the primary explanation of the low ρ/θ ratio. A similar pattern was shown for the relative effect (r/m) of recombination and mutation (1.690 for the entire population and 1.784 for M5S2). Since ClonalFrameML identifies recombination events based on the clustered distribution of single nucleotide variants, which are primarily introduced by recombination with external lineages (i.e., those evolutionarily separated from the members under study) (36, 37), the appreciable difference between ρ/θ and r/m shown here suggests that such recombination plays an important role in the genetic differentiation of the sampled CHUG lineage.

At first glance, the results from ClonalFrameML and PopCOGenT appear to be contradictory. On one hand, both the recombination rate (ρ/θ) and relative effect (r/m), as measured by ClonalFrameML, are largely invariable between and within the populations defined by PopCOGenT. On the other hand, recombination must be more frequent within than between populations, as this is the criterion used by PopCOGenT to delineate population boundaries. To this end, it is worth emphasizing that PopCOGenT exclusively counts recombination events occurring within the studied lineage because it detects gene flow based on the enrichment of identical genomic regions, whereas ClonalFrameML mainly focuses on recombination with external lineages. Therefore, the analyses by ClonalFrameML and PopCOGenT are complementary.

The earliest speciation event has the most profound impacts on genomic microdiversity. We provide three lines of evidence to support the hypothesis that the first split, separating M1M2 from M3M4M5, left the strongest signatures in the genomes. F_{ST} is a proxy for the levels of allelic fixation and genetic differentiation in core genomes (38, 39), and the core DNA is considered to have reached a high level of differentiation if the F_{ST} value is statistically significant ($P < 0.05$) and is greater than 0.25 (40, 41). According to this criterion, 82.9% of the 1,788 single-copy core genes shared by M1M2 and M3M4M5 (Fig. 1B) reach a high level of differentiation, which is significantly greater (two-proportion z-test, $P < 0.001$) than that following the split of M3M4 from M5 (78.8%) (Fig. 1C) and that of the ongoing split of M5S1 and M5S2 (66.8%) (Fig. 1D). F_{ST} cannot be calculated for other splits, including the genetic separation of M1 from M2 or that of M3 from M4, because one of the populations under comparison (M1 in the former comparison and M3 in the latter) is comprised of only a single member.

An important evolutionary mechanism driving population differentiation in core genomic regions is novel allelic replacements via homologous recombination with distant relatives. Such replacements are expected to leave genetic signatures at synonymous sites in protein-coding genes, which manifest as unusually large synonymous substitution rates (d_s) at the recombined loci compared to those of the remaining genes (42). The underlying principle is that selective forces on nucleotide substitutions at synonymous sites are weak. Thus, unusually high d_s , resulting from replacements via homologous recombination with divergent lineages, in a given gene are unlikely to be overwritten by changes induced by other processes, such as mutation and selection. In the present study, this approach is used to compare the roles of sequential speciation events in shaping the genomic diversity of the 33 CHUG members. In total, we identified 38 core gene families showing unusually large d_s values (see Text S1.6 for more

FIG 1 Legend (Continued)

populations from M1 to M5 as well as the two subpopulations, M5S1 and M5S2, within M5 that were defined by PopCOGenT are mapped to the maximum likelihood phylogenomic tree constructed by IQ-TREE, with the root position determined by the MAD algorithm. Solid circles in the phylogeny indicate nodes with bootstrap values of 100% in the 1,000 bootstrapped replicates. Three split events discussed in the main paper are marked with arrows. (B to D) The distribution of the F_{ST} values across 1,788 core genes between the lineages, resulting from each of the three split events. Gene families showing $F_{ST} > 0.25$ with statistical significance ($P < 0.05$) are marked in red, and others are marked in gray. For the dot plots (top), genes are ordered along the closed genome of HKCCA1288. For the histogram (bottom), the count represents the number of gene families. (E) The dendrogram is constructed based on gene presence and absence using the complete linkage method implemented in the R package, “*phreatmap*”.

details; Data set S1b and Data set S1c). By integrating gene tree topological structures and the pairwise d_s values among gene members within a family, we were able to infer at which phylogenetic branches the divergent alleles were introduced to the population. For example, 25 of the 38 gene families each consistently show unusually large d_s values in pairwise comparisons between M1M2 and M3M4M5 but generally small d_s values when compared to members within either M1M2 or M3M4M5, indicating that allelic replacements occurred at the last common ancestor (LCA) of either M1M2 or M3M4M5 (Fig. S2). Likewise, we inferred that five gene families were replaced at the LCA of either M3M4 or M5, and seven gene families were replaced at the LCA of either M5S1 or M5S2 (Fig. S2). Note that 15 gene families were subjected to multiple allelic replacements at different evolutionary branches and may have contributed to several speciation events (Fig. S2). Therefore, this analysis provides new evidence in support of the idea that the earliest split event left the strongest imprints on the diversity of the core genome.

Further evidence comes from the accessory genome. Consistent with the phylogenomic tree, the topology of the gene content dendrogram, constructed based on gene presence and absence in the 33 CHUG genomes (Fig. 1E), supports the separation of M1M2 from M3M4M5. Within M3M4M5, however, the dendrogram shows that the subpopulation M5S2 clusters with M3M4 instead of the subpopulation M5S1, suggesting that genetic separation within M3M4M5 has not been completed at the accessory genome. Therefore, this analysis supports the claim that the genomic diversity of the 33 CHUG members was most profoundly impacted by the earliest split event. Over 60% (2,108 out of 3,451) of the accessory gene families are mapped to genomic islands (GI), and a dendrogram based on these GI-associated genes still supports both the split between M1M2 and M3M4M5 and a mixed relationship within M3M4M5 (Fig. S3). This suggests that GIs may contribute appreciably to the population differentiation of these CHUG populations. Pseudogenes and mobile genetic elements other than GIs, such as insertion sequences and prophages, are also abundant (Data set S1a; Fig. 2), suggesting that they are also important drivers of the genomic heterogeneity of the sampled CHUG members. An excess of pseudogenes, GIs, and other mobile elements makes CHUG unique compared to most known marine bacterioplankton lineages with reduced genomes, in which these elements are depleted (43–45).

Differential utilization of carbon sources correlates with the earliest speciation event. Since population-specific gene families (exclusively and ubiquitously found in one population but absent in another) may provide clues regarding functional differentiation, we focused on the 210 M1M2-specific and 115 M3M4M5-specific gene families defined by Roary (Data Set S1d and e), which include 152 and 91 families, respectively, that can be mapped to Clusters of Orthologous Groups (COG) families with known functions. Of these, 33 M1M2-specific and 23 M3M4M5-specific families defined by Roary are potentially involved in the utilization of a variety of carbon sources (mostly carbohydrates), suggesting that the availability of distinct carbon sources and the differential utilization of these carbon sources may have contributed to the earliest speciation event. It is worth mentioning that the differential utilization of carbohydrates may also contribute to follow-on speciation events, since carbohydrate utilization genes are part of the M3M4-specific (2 out of 23), M5-specific (2 out of 4), M5S1-specific (6 out of 32), and M5S2-specific (7 out of 21) gene families (Fig. 2; Data Set S1f–i).

We predicted 20 carbon sources that show genetic differentiation between M1M2 and M3M4M5 (Fig. 2; Table 1; Text S2.1). To systematically characterize the phenotypic differentiation involved in organic substrate utilization, we employed BiOLOG phenotype microarrays, which consist of 190 distinct carbon sources. There are two benefits of using this high-throughput technique. First, it allows testing the substrates underlying the genetic differentiation, since it includes 16 of the 20 above-mentioned carbon sources (Table 1). Second, it allows testing additional substrates that are not predicted by a bioinformatic analysis. One caveat of using this technique is that BiOLOG plates contain all carbon substrates at the same concentration, but bacterial responses may depend on the concentrations of the carbon compounds. That is, negative BiOLOG

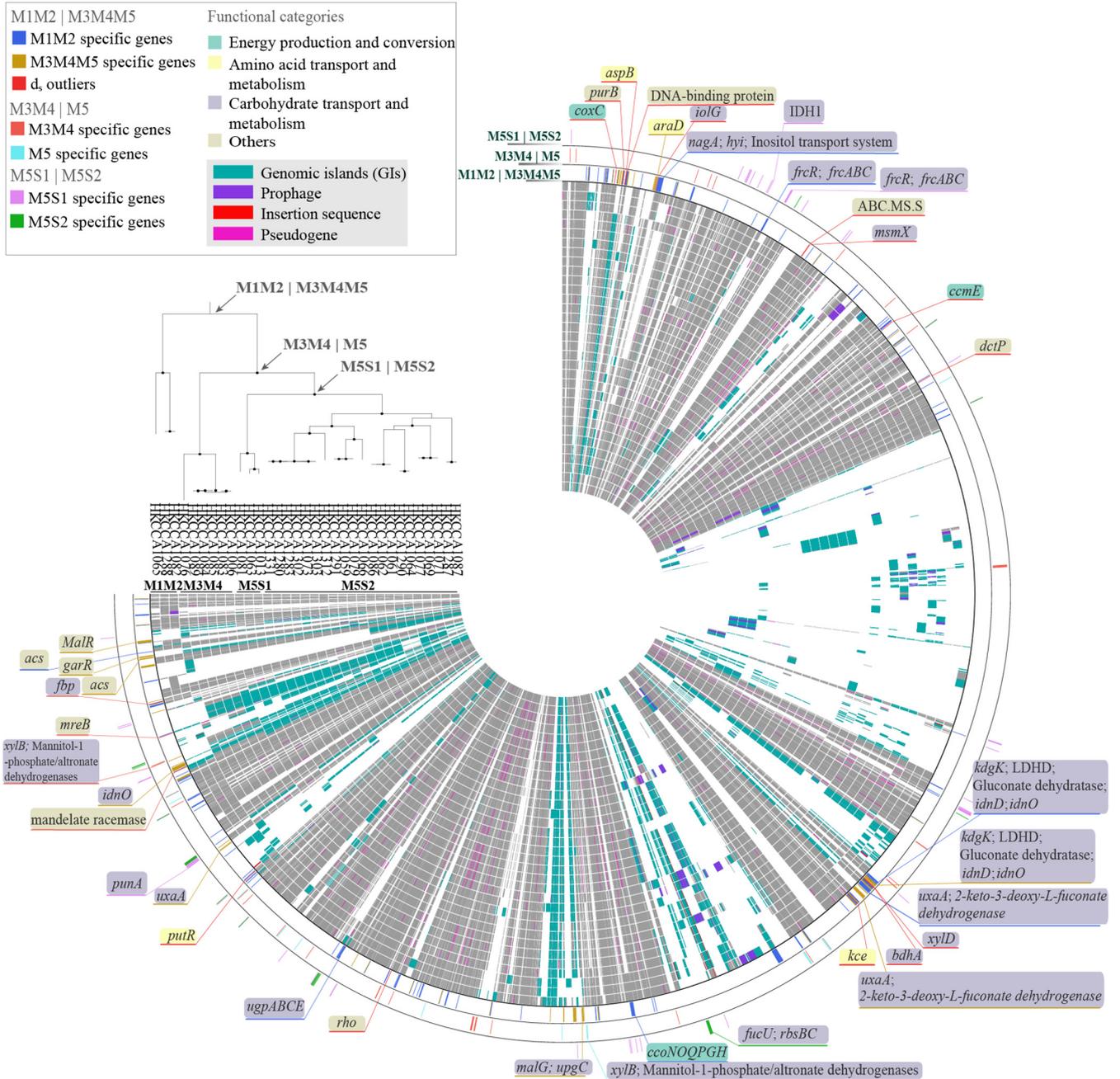


FIG 2 The pangenome of the 33 CHUG isolates. All of the orthologous gene families identified by Roary are positioned according to the closed genome, HKCCA1288. From the inner to the outer circle: (1 to 33) the genomes of the 33 CHUG isolates, arranged in line with their phylogenomic tree, which is shown in the top-left corner. Genomic islands, prophages, insertion sequences, and pseudogenes are marked with different colors and are mapped to each genome; (34) M1M2-specific genes, M3M4M5-specific genes, and core genes showing unusually large d_5 values are marked with different colors and are placed according to the coordinates of the HKCCA1288 genome; (35) M3M4-specific genes and M5-specific genes; (36) M5S1-specific genes and M5S2-specific genes. Gene families falling in the functional categories, “energy production and conversion”, “amino acid transport and metabolism”, and “carbohydrate transport and metabolism”, are each attached with a gene name and framed in a box with a background color corresponding to a functional category. Each box is connected with a line that is colored according to whether the gene family is part of the population-specific accessory genes or the d_5 outlier core genes.

results may turn positive in a concentration different from that under BiOLOG and vice versa.

Of the 16 carbon sources mentioned above, 9 are potentially made by macroalgae (Table 1). Three macroalgae-derived compounds (β -hydroxybutyric [46], L-fucose [27–29], and acetic acid [47]) each were differentially utilized by M1M2 and M3M4M5 (Fig. 3; Table 1), which is correlated with the genetic differentiation of the corresponding gene

TABLE 1 Organic compounds that support genetic or phenotypic differentiation between M1M2 and M3M4M5^a

Compounds	M1M2-specific ^b	M3M4M5-specific ^b	<i>d_s</i> outliers ^c	Different Roary families with same COG annotation ^d	BiOLOG assay	Macroalgae-derived
Hydroxypyruvate	Yes				-	
N-acetylglucosamine	Yes				No growth	
D-fructose	Yes				Differential growth ^e	Brown, red, and green (51)
Myo-inositol	Yes		Yes		No growth	Brown, red, and green (51)
Glycerol-3-phosphate	Yes			Yes	No growth	Red (89)
D-arabinose		Yes			Equal growth	
Maltose		Yes			No growth	Green (90)
D-glycerate		Yes			-	
Tartronate semialdehyde		Yes			-	
β -hydroxybutyric acid			Yes		Differential growth	Brown (46)
D-xylose			Yes		Equal growth	Brown, red, and green (50)
D-galacturonate				Yes	Equal growth	
D-glucuronate				Yes	No growth except one replicate of the M2 representative	Brown (52)
Methyl D-Lactate				Yes	No growth except one replicate of the M2 representative	
Pyruvate				Yes	Equal growth	
D-gluconate				Yes	No growth	
Ketogluconate				Yes	Equal growth	
2-keto-3-deoxy-L-fuconate				Yes	-	
L-fucose				Yes	Differential growth	Brown (27 to 29)
Acetic acid				Yes	Differential growth	Red (47)
Glycolic acid					Differential growth	Brown (49)
Sodium formate					Differential growth	
Mono-methyl succinate					Differential growth	
D-arabitol					Differential growth	
D-fucose					Differential growth	
γ -hydroxybutyric Acid					Differential growth	

^aCompounds potentially made by macroalgae are marked in the last column.

^bThe “M1M2-specific” and “M3M4M5-specific” gene families are exclusively present in M1M2 and M3M4M5, respectively.

^cThe “*d_s* outlier” gene families are shared by all members of M1M2 and M3M4M5 but were subjected to novel allelic replacement.

^dGene families linked to “different Roary gene families with the same COG annotation” each consist of two families identified by Roary but are annotated with the same function. The gene locus, COG id, gene names, and functions of the involved gene families can be found in Data Set S1j.

^eThe differential growth is between M1 and the remaining populations. In other cases, differential growth was observed between M1M2 and M3M4M5.

families (Data set S1j; Text S2.1). Of these compounds, L-fucose is a major component of macroalgal fucoidan and constitutes up to 40% of the monosaccharides in some brown algae (27). Acetic acid can be released by some red algae (47), though it is commonly produced by other marine organisms, including bacteria (48). Moreover, M1M2 showed better growth than did M3M4M5 when utilizing another six different carbon sources in the phenotype microarray assay (Fig. 3B), including glycolic acid, which is a major component in brown algae (49). While gene differentiation between M1M2 and M3M4M5 was not detected for these six compounds, genetic divergence at intergenic regions for regulation cannot be ruled out. It is also worthy of note that 11 other carbon sources led to comparable respiratory intensity in M1M2 and M3M4M5. Some of these carbon sources may be derived from macroalgae, such as D-xylose (brown, red, and green macroalgae) (50), D-ribose (brown and green macroalgae) (51), and D-arabinose (brown macroalgae) (51, 52). Our results suggest that while M1M2 and M3M4M5 are each optimized for utilizing a specific group of carbon sources, the sampled CHUG members might be, in general, well-prepared to make use of the resources available in macroalgal ecosystems (Fig. 3; Fig. S4 and S5).

The divergences and coherences between M1M2 and M3M4M5 members at the phenotypic level were further analyzed by clustering the tested strains based on respiration levels as a proxy for growth responses to carbon source availability. The resulting dendrogram (Fig. 3A) shows that the M1M2 members are well-separated from the M3M4M5 members, whereas clades within M3M4M5 are mixed. This evidence further supports the idea that the earliest speciation event, that is, the event splitting M1M2

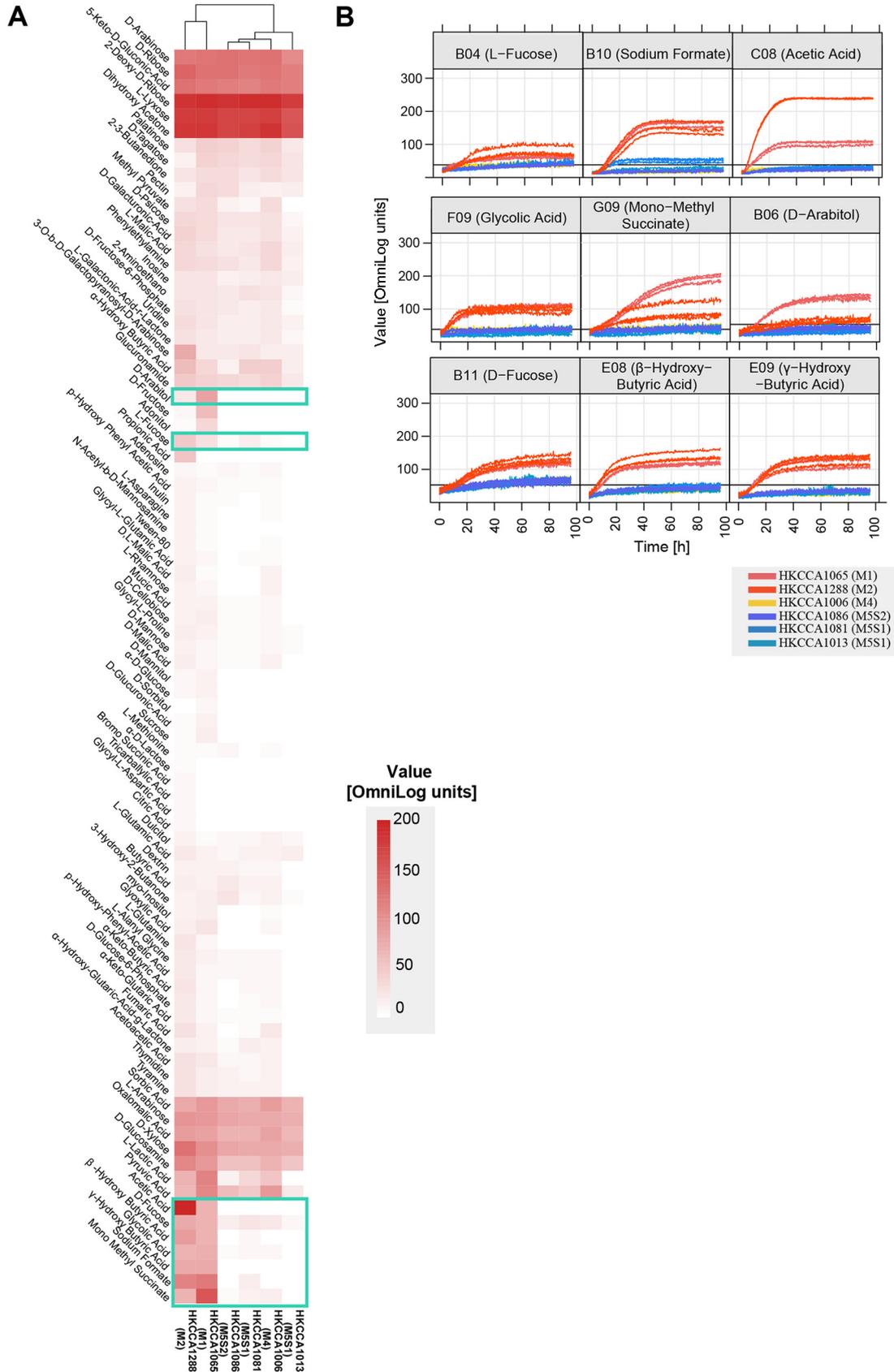


FIG 3 (A) The heat map of the respiration values of a few representative strains of the populations defined by PopCOGenT at
(Continued on next page)

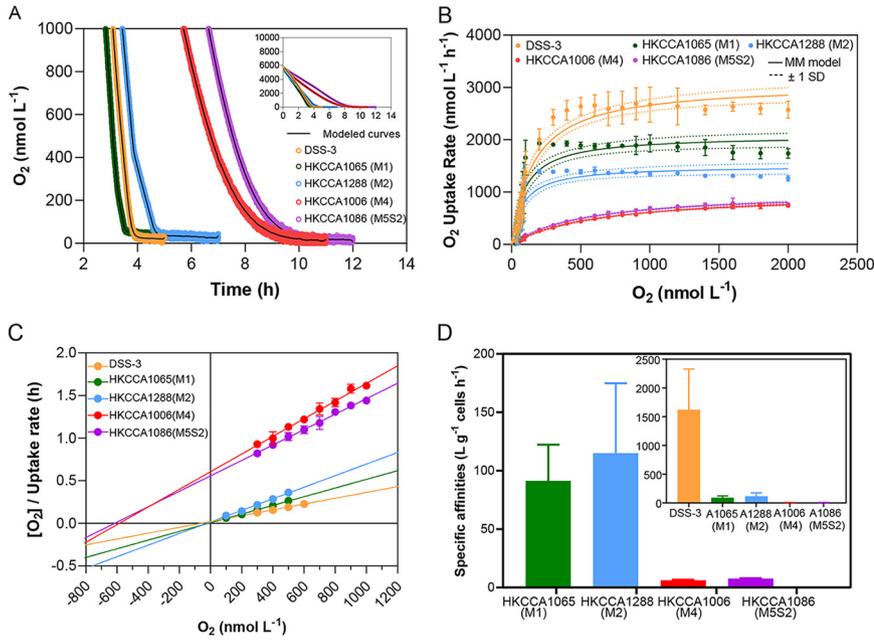


FIG 4 Oxygen uptake measurements. (A) Oxygen consumption of two CHUG isolates (HKCCA1288 and HKCCA1065) from M1M2, two CHUG isolates (HKCCA1006 and HKCCA1086) from M3M4M5, and *Ruegeria pomeroyi* DSS-3 under microaerobiosis. (B) The relationship between oxygen concentration and oxygen uptake rates at an oxygen level of <2 μmol/L. The solid line is a model fitted based on the Michaelis-Menten equation, and the nearby dotted line shows the standard deviation estimated by this equation. The error bars of the filled circles represent the standard deviations of three replicates. (C) The Hanes-Woolf plot of oxygen uptake as a function of oxygen concentration, ranging from 100 nmol/L to 1 μmol/L. The error bars of the filled circles represent the standard deviations of three replicates. (D) The estimated specific affinity of oxygen for the four CHUG isolates and *R. pomeroyi* DSS-3.

and M3M4M5, has shaped the physiological diversity in the 33 CHUG members, whereas the follow-on speciation events have made a limited contribution.

Differential adaptation to low and high oxygen niches is another key driver of the earliest speciation event. Another notable observation is that while all CHUG members possess the cytochrome aa₃ oxidase (*ctaBCDEG*), which is known to have a low oxygen (O₂) affinity and thereby enables them to respire under high O₂ conditions (53), all M1M2 members additionally and exclusively harbor the cytochrome cbb₃ oxidase (*ccoGHNOPQ*) (Fig. 2), which is known to exhibit a high O₂ affinity that may function under microaerobic conditions (54–56). Moreover, a regulatory gene, *FnrL*, is located upstream of the *ccoGHNOPQ* operon, which may regulate *cbb₃* oxidase gene expression (57, 58) when M1M2 members switch from aerobic to microaerobic niches.

Our O₂ uptake experiments revealed that, under microaerobic conditions, O₂ was consumed slower in the M3M4M5 cultures than in both the M1M2 cultures and in *Ruegeria pomeroyi* DSS-3 (Fig. 4A), a model Roseobacter with a large genome (4.6 Mbp) that harbors both the aa₃ oxidase and the cbb₃ oxidase (59). Kinetic parameters, including the maximum respiration rate (*V*_{max}) and the half-saturation constants (apparent *K*_m values) are used to describe the respiratory rates as a function of O₂ concentration using a Michaelis-Menten model (Fig. 4B and C). Estimated *K*_m values for M1M2 and *R. pomeroyi* DSS-3 were found to be lower than those of M3M4M5 (Data Set S1k), which is consistent with the fact that only M1M2 and *R. pomeroyi* DSS-3 possess the cbb₃ oxidase with a high O₂ affinity. Next, the specific affinity for O₂, which quantifies

FIG 3 Legend (Continued)

96 h, measured by BiOLOG. The clustering is constructed using all 190 carbon sources provided by the two microplates, PM01 and PM02, but only those 85 substrates that showed a respiration value over 10 OmniLog units in at least one strain are displayed. Substrates differentially utilized by M1M2 and M3M4M5 are framed with green boxes. (B) The respiration curves of the representative strains fed with the framed substrates in (A).

the ability of bacteria to take up O₂ at low concentrations (60, 61), shows an order-of-magnitude difference between M1M2 and M3M4M5 (Fig. 4D; Data Set S1k), further indicating their functional divergence through microaerobiosis.

Macroalgae release adequate bioavailable DOCs to enable the proliferation of ambient heterotrophic microorganisms (62–64), which subsequently deoxygenate the ambient waters at a microspatial scale (65). Moreover, the DOC release process is dynamic so that the spatial and temporal depletion of O₂ may exist in macroalgal ecosystems (64, 66), and this may contribute to the formation of the two main ecotypes of CHUG: M1M2 and M3M4M5. A further observation is that *R. pomeroyi* DSS-3 shows a much larger specific affinity than does CHUG, which implies its greater potential to acquire O₂ from exceedingly low O₂ habitats (Fig. 4D; Data set S1k). This appears to be an adaptive strategy of this Roseobacter; since it is often associated with organic particles in the water column and in the interior of the particles, O₂ may be depleted owing to aerobic degradation at the outer part of the particles (67, 68).

Concluding remarks and caveats. CHUG is a unique Roseobacter lineage that differs from all other pelagic Roseobacter members in that its members have some of the smallest genomes in the Roseobacter group, their global distribution is not correlated with phytoplankton abundance, they lack the ability to de novo synthesize vitamin B₁₂, and they cannot grow on many carbon compounds made by phytoplankton (21). The decoupling from the common habitat in the pelagic ocean, namely, phytoplankton, shared by most pelagic Roseobacter members, prompted us to explore potential habitats for CHUG. Here, we provided evidence that the speciation of the CHUG members correlates with the differential utilization of low-molecular-weight labile organic compounds and the differential ability to explore low-oxygen niches. Since these resources and conditions are transient in the water column, our findings suggest that CHUG may explore ephemeral habitats where nutrient concentrations may be occasionally high but oxygen levels may be transiently low. The available data are consistent with the idea that brown algae are likely an important habitat of the CHUG members, since they were sampled from the ambient seawater of the brown algae *Sargassum* and since they can grow on L-fucose, the dominant carbohydrate found in brown algae, as the sole carbon source, whereas all other pelagic Roseobacter lineages (21), such as DC5-80-3 (also called “RCA”), CHAB-I-5, and NAC11-7, lack the key gene (*fucA*) that enables the use of this compound. Nevertheless, other possible habitats cannot be ruled out. For example, fresh organic aggregates may also enrich nutrients and deplete oxygen in their ambient water immediately surrounding the particles, and there is no compelling reason that L-fucose cannot be an important component in such habitats. Given the seasonal occurrence of *Sargassum*, future researchers may perform longitudinal sampling and combine culture-based approaches with metagenomics and meta-transcriptomics to validate brown algae as an important habitat for CHUG.

MATERIALS AND METHODS

The 33 CHUG isolates were obtained from a single 50 mL sample of water surrounding brown alga *Sargassum hemiphyllum* at Lobster Bay (22.309° N, 114.3015° E), Hong Kong SAR, China, on 15 Mar 2017. All CHUG members were isolated using a modified marine basal medium (MBM) (14) and cultivated with 2216 marine agar (BD Difco, USA). The 33 CHUG genomes were sequenced using the BGISEQ-500 platform in PE100 sequencing mode. Genomes were assembled using SPAdes v3.9.1 (69) with high quality reads, and the quality of assemblies was determined with CheckM v1.0.7 (Data set S1a) (70). Genes were predicted with Prokka v1.11 (71) and annotated with the RAST server v2.0 (72), the CDD database v3.16 (73), the KEGG database v82.0 (74, 75), and the COG database (76). Genomic islands, prophages, insertion sequences (IS), and pseudogenes were identified using the online IslandViewer 4 server (77), PHASTER web server (78), ISfinder web server (79), and a modified procedure of Psi-Phi (14, 80), respectively.

To construct the clonal evolutionary history of the CHUG members, a maximum likelihood (ML) phylogenomic tree was constructed using IQ-TREE v1.6.5 (81) with a shared genomic DNA alignment, in which all recombinant sites were masked by Gubbins 2.4.1 (82). Note that the shared genomic DNA alignment was derived from a whole-genome alignment and includes both protein-coding genes and non-coding genomic regions. To root the tree, we initially tried the outgroup-based method. However, the most closely related sister lineage among all available genome-sequenced lineages is connected with the CHUG lineage through a long branch (Fig. S6A), suggesting that it is not an appropriate outgroup and that using it to root the CHUG phylogeny may bias the placement of the root. Long branches may

result from using fast evolving sites, such as synonymous sites, in protein-coding genes and noncoding DNA. This hypothesis was validated with a phylogenomic tree construction based on the concatenated amino acid alignment of the core genes shared by the sister group and the CHUG lineage (Fig. S6A, S6B). In this amino acid sequence-based tree, however, the branch connecting CHUG and its sister group is still long compared to the branches within the CHUG lineage (Fig. S6B). We therefore employed an outgroup-free method, the minimal ancestor deviation (MAD) approach, to root the phylogenomic tree (Fig. S6C). This method considers each branch as a possible root position and infers the ancestor-descendant relationships of all possible ancestral nodes in the tree. For each relationship, the mean relative deviation from the clock-likeness is evaluated, and the branch with the minimal relative deviation is considered the one containing the root node. In the MAD analysis, the sister group was removed, and only the 33 CHUG genomes were used. We chose the phylogenomic tree based on the shared genomic DNA for MAD rooting because: (i) given the close relationship among the CHUG members, nucleotide sequences have higher resolution, and nucleotide substitution is less likely to reach saturation; and (ii) only nucleotide sequences allow Gubbins to detect and mask the recombined regions, an essential step in constructing a clonal evolutionary relationship for the CHUG lineage (82).

Next, genetically isolated populations were defined with PopCOGenT (83). The rationale of this method was that since bacterial speciation often proceeds rapidly, the delineation of population boundaries should be based on gene flow events that occurred recently (83). One main advantage of PopCOGenT over earlier methods, such as ConSpecFix (84), is that it allows separating recent transfer events from historical ones (83). Ancestral nodes in the phylogenomic tree that lead to speciation events can be easily identified based on the population membership defined by PopCOGenT. Further evidence for population differentiation was provided at the core genome by computing the fixation index (F_{ST}) between delineated populations using Arlequin v3.5 (85) and at the accessory genome by clustering the genomes based on the gene presence and absence using the complete linkage method implemented in the R package, “*pheatmap*” (86). Prior studies demonstrated that population differentiation in members of the marine *Roseobacter* group is often achieved by replacing alleles at some core genes with novel and potentially adaptive copies derived from external species (8). This mechanism was explored using the synonymous substitution rate (d_s) clustering approach, which makes pairwise comparisons across single-copy orthologous genes and across all genomes under comparison (42). The rationale is that synonymous substitution is largely neutral, and thus, genes with unusually large d_s values likely result from recombination-driven novel allele replacements. This approach helps to infer at which phylogenetic branches the divergent alleles were introduced into the population, so it was implemented here to determine the relative importance of allele replacement in core genes in the successive speciation events.

Since the CHUG populations differ in genes involved in carbon source utilization and oxygen respiration, we conducted physiological assays of substrate utilization through BiOLOG Phenotype Microarrays (PM) (87) and through measurements of oxygen uptake under microaerobic conditions using optical oxygen sensors (88), respectively. For the former, PM01 and PM02, which contain 190 carbon sources, were used. In addition to checking the differential utilization of the carbon sources predicted by the genetic differences, we also constructed a heat map based on the maximal respiration intensity at 96 h by using the complete linkage method implemented in the R package, “*pheatmap*” (86), to provide an overall representation of the substrate utilization among populations. For the latter, the kinetic parameters describing the oxygen uptake rates, including the maximum uptake rate (V_{max}) and the apparent half-saturation constant (K_m), were estimated. Specific affinities to oxygen were calculated following a method described by Button et al. (60). Additional methodological details are provided in Supplemental Methods in Text S1.

Data availability. Raw reads and the assembled genomic sequences of the new CHUG genomes are available in the NCBI GenBank database under the accession number [PRJNA770146](https://doi.org/10.1093/bioinformatics/btad014).

Code availability. The scripts used for the phylogenetic tree construction, F_{ST} calculation, and Biolog data processing are available from an online repository: <https://github.com/luolab-cuhk/CHUG-pop-microdiversity>.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLSX file, 0.2 MB.

TEXT S1, DOCX file, 0.1 MB.

TEXT S2, DOCX file, 0.03 MB.

FIG S1, PDF file, 1.2 MB.

FIG S2, PDF file, 0.2 MB.

FIG S3, PDF file, 0.9 MB.

FIG S4, PDF file, 2.7 MB.

FIG S5, PDF file, 2.6 MB.

FIG S6, PDF file, 0.2 MB.

ACKNOWLEDGMENTS

We thank Hiu Ning Leung, Ho Leung Tsang, Tsz Yan Ng, and Kwan Ting Wong for helping with the sample collection, as well as Weiyun Guo and Siyao Li for assisting

with the bacterial isolation. We also thank Tianhua Liao, Sishuo Wang, and Ying Sun for their assistance with the bioinformatics analysis. This research was supported by the Shenzhen Science and Technology Committee (JCYJ20180508161811899), the National Science Foundation of China (41776129), the Hong Kong Research Grants Council General Research Fund (14163917), the Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16), and the Direct Grant of CUHK (4053257 and 3132809).

REFERENCES

- Larkin AA, Martiny AC. 2017. Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ Microbiol Rep* 9:55–70. <https://doi.org/10.1111/1758-2229.12523>.
- Fodelianakis S, Washburne AD, Bourquin M, Pramateftaki P, Kohler TJ, Styllas M, Tolosano M, De Staercke V, Schön M, Busi SB, Brandani J, Wilmes P, Peter H, Battin TJ. 2022. Microdiversity characterizes prevalent phylogenetic clades in the glacier-fed stream microbiome. 3. *ISME J* 16:666–675. <https://doi.org/10.1038/s41396-021-01106-6>.
- Needham DM, Sachdeva R, Fuhrman JA. 2017. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J* 11:1614–1629. <https://doi.org/10.1038/ismej.2017.29>.
- García-García N, Tamames J, Linz AM, Pedrós-Alió C, Puente-Sánchez F. 2019. Microdiversity ensures the maintenance of functional microbial communities under changing environmental conditions. *ISME J* 13:2969–2983. <https://doi.org/10.1038/s41396-019-0487-8>.
- Xu Q, Luo G, Guo J, Xiao Y, Zhang F, Guo S, Ling N, Shen Q. 2022. Microbial generalist or specialist: intraspecific variation and dormancy potential matter. *Mol Ecol* 31:161–173. <https://doi.org/10.1111/mec.16217>.
- Chase AB, Karaoz U, Brodie EL, Gomez-Lunar Z, Martiny AC, Martiny JBH. 2017. Microdiversity of an abundant terrestrial bacterium encompasses extensive variation in ecologically relevant traits. *mBio* 8:e01809-17. <https://doi.org/10.1128/mBio.01809-17>.
- Ahlgren NA, Perelman JN, Yeh Y-C, Fuhrman JA. 2019. Multi-year dynamics of fine-scale marine cyanobacterial populations are more strongly explained by phage interactions than abiotic, bottom-up factors. *Environ Microbiol* 21:2948–2963. <https://doi.org/10.1111/1462-2920.14687>.
- Wang X, Zhang Y, Ren M, Xia T, Chu X, Liu C, Lin X, Huang Y, Chen Z, Yan A, Luo H. 2020. Cryptic speciation of a pelagic *Roseobacter* population varying at a few thousand nucleotide sites. *ISME J* 14:3106–3119. <https://doi.org/10.1038/s41396-020-00743-7>.
- Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, Polz MF, Alm EJ. 2012. Population genomics of early events in the ecological differentiation of bacteria. *Science* 336:48–51. <https://doi.org/10.1126/science.1218198>.
- Bendall ML, Stevens SLR, Chan LK, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B, Froula J, Kang D, Tringe SG, Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J* 10:1589–1601. <https://doi.org/10.1038/ismej.2015.241>.
- Shapiro BJ, Polz MF. 2015. Microbial speciation. *Cold Spring Harb Perspect Biol* 7:a018143. <https://doi.org/10.1101/cshperspect.a018143>.
- Hehemann J-H, Arevalo P, Datta MS, Yu X, Corzett CH, Henschel A, Preheim SP, Timberlake S, Alm EJ, Polz MF. 2016. Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. *Nat Commun* 7:1–10. <https://doi.org/10.1038/ncomms12860>.
- Bruto M, Labreuche Y, James A, Piel D, Chenivresse S, Petton B, Polz MF, Le Roux F. 2018. Ancestral gene acquisition as the key to virulence potential in environmental *Vibrio* populations. *ISME J* 12:2954–2966. <https://doi.org/10.1038/s41396-018-0245-3>.
- Chu X, Li S, Wang S, Luo D, Luo H. 2021. Gene loss through pseudogenization contributes to the ecological diversification of a generalist *Roseobacter* lineage. *ISME J* 15:489–502. <https://doi.org/10.1038/s41396-020-00790-0>.
- Luo D, Wang X, Feng X, Tian M, Wang S, Tang S-L, Ang P, Yan A, Luo H. 2021. Population differentiation of *Rhodobacteraceae* along with coral compartments. *ISME J* 15:3286–3302. <https://doi.org/10.1038/s41396-021-01009-6>.
- Lin X, McNichol J, Chu X, Qian Y, Luo H. 2022. Cryptic niche differentiation of novel sediment ecotypes of *Ruegeria pomeroyi* correlates with nitrate respiration. *Environ Microbiol* 24:390–403. <https://doi.org/10.1111/1462-2920.15882>.
- Pereira-Flores E, Glöckner FO, Fernandez-Guerra A. 2019. Fast and accurate average genome size and 16S rRNA gene average copy number computation in metagenomic data. *BMC Bioinform* 20:453. <https://doi.org/10.1186/s12859-019-3031-y>.
- Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, Ding H, Marttinen P, Malmstrom RR, Stocker R, Follows MJ, Stepanauskas R, Chisholm SW. 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344:416–420. <https://doi.org/10.1126/science.1248575>.
- Clingenpeel S, Clum A, Schwientek P, Rinke C, Woyke T. 2014. Reconstructing each cell's genome within complex microbial communities—dream or reality? *Front Microbiol* 5:771. <https://doi.org/10.3389/fmicb.2014.00771>.
- Chen Z, Wang X, Song Y, Zeng Q, Zhang Y, Luo H. 2022. *Prochlorococcus* have low global mutation rate and small effective population size. *Nat Ecol Evol* 6:183–194. <https://doi.org/10.1038/s41559-021-01591-0>.
- Feng X, Chu X, Qian Y, Henson MW, Lanclos VC, Qin F, Barnes S, Zhao Y, Thrash JC, Luo H. 2021. Mechanisms driving genome reduction of a novel *Roseobacter* lineage. *ISME J* 15:3576–3586. <https://doi.org/10.1038/s41396-021-01036-3>.
- Luo H, Moran MA. 2014. Evolutionary ecology of the marine *Roseobacter* clade. *Microbiol Mol Biol Rev* 78:573–587. <https://doi.org/10.1128/MMBR.00020-14>.
- Seymour JR, Amin SA, Raina J-B, Stocker R. 2017. Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nat Microbiol* 2:1–12. <https://doi.org/10.1038/nmicrobiol.2017.65>.
- Cooper MB, Kazamia E, Helliwell KE, Kudahl UJ, Sayer A, Wheeler GL, Smith AG. 2019. Cross-exchange of B-vitamins underpins a mutualistic interaction between *Ostreococcus tauri* and *Dinoroseobacter shibae*. *ISME J* 13:334–345. <https://doi.org/10.1038/s41396-018-0274-y>.
- Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ, Dearth SP, Mooy BASV, Campagna SR, Kujawinski EB, Armbrust EV, Moran MA. 2015. Cryptic carbon and sulfur cycling between surface ocean plankton. *Proc Natl Acad Sci U S A* 112:453–457. <https://doi.org/10.1073/pnas.1413137112>.
- Wagner-Döbler I, Ballhausen B, Berger M, Brinkhoff T, Buchholz I, Bunk B, Cypionka H, Daniel R, Drepper T, Gerdts G, Hahnke S, Han C, Jahn D, Kalhoefer D, Kiss H, Klenk H-P, Kyrpides N, Liebl W, Liesegang H, Meincke L, Pati A, Petersen J, Piekarski T, Pommerenke C, Pradella S, Pukall R, Rabus R, Stackebrandt E, Thole S, Thompson L, Tielens P, Tomasch J, von Jan M, Wanphrut N, Wichels A, Zech H, Simon M. 2010. The complete genome sequence of the algal symbiont *Dinoroseobacter shibae*: a hitchhiker's guide to life in the sea. *ISME J* 4:61–77. <https://doi.org/10.1038/ismej.2009.94>.
- Sellimi S, Kadri N, Barragan-Montero V, Laouer H, Hajji M, Nasri M. 2014. Fucans from a Tunisian brown seaweed *Cystoseira barbata*: structural characteristics and antioxidant activity. *Int J Biol Macromol* 66:281–288. <https://doi.org/10.1016/j.ijbiomac.2014.02.041>.
- Rioux L-E, Turgeon SL, Beaulieu M. 2009. Effect of season on the composition of bioactive polysaccharides from the brown seaweed *Saccharina longicuris*. *Phytochemistry* 70:1069–1075. <https://doi.org/10.1016/j.phytochem.2009.04.020>.
- Ale MT, Mikkelsen JD, Meyer AS. 2011. Important determinants for fucoidan bioactivity: a critical review of structure-function relations and extraction methods for fucose-containing sulfated polysaccharides from brown seaweeds. *Mar Drugs* 9:2106–2130. <https://doi.org/10.3390/md9102106>.
- Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R. 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 12:635–645. <https://doi.org/10.1038/nrmicro3330>.

31. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* 102:2567–2572. <https://doi.org/10.1073/pnas.0409727102>.
32. Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, Krause DJ, Whitaker RJ. 2012. Patterns of gene flow define species of thermophilic archaea. *PLoS Biol* 10:e1001265. <https://doi.org/10.1371/journal.pbio.1001265>.
33. Wielgoss S, Didelot X, Chaudhuri RR, Liu X, Weedall GD, Velicer GJ, Vos M. 2016. A barrier to homologous recombination between sympatric strains of the cooperative soil bacterium *Myxococcus xanthus*. *ISME J* 10:2468–2477. <https://doi.org/10.1038/ismej.2016.34>.
34. Kopac S, Wang Z, Wiedenbeck J, Sherry J, Wu M, Cohan FM. 2014. Genomic heterogeneity and ecological speciation within one subspecies of *Bacillus subtilis*. *Appl Environ Microbiol* 80:4842–4853. <https://doi.org/10.1128/AEM.00576-14>.
35. Choudoir MJ, Buckley DH. 2018. Phylogenetic conservatism of thermal traits explains dispersal limitation and genomic differentiation of *Streptomyces* sister-taxa. *ISME J* 12:2176–2186. <https://doi.org/10.1038/s41396-018-0180-3>.
36. Didelot X, Wilson DJ. 2015. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* 11:e1004041. <https://doi.org/10.1371/journal.pcbi.1004041>.
37. Arnold BJ, Huang I-T, Hanage WP. 2022. Horizontal gene transfer and adaptive evolution in bacteria. *Nat Rev Microbiol* 20:206–213. <https://doi.org/10.1038/s41579-021-00650-4>.
38. Excoffier L, Hofer T, Foll M. 2009. Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb)* 103:285–298. <https://doi.org/10.1038/hdy.2009.74>.
39. Via S. 2009. Natural selection in action during speciation. *Proc Natl Acad Sci U S A* 106:9939–9946. <https://doi.org/10.1073/pnas.0901397106>.
40. Wright S. 1984. Evolution and the genetics of populations, volume 4: variability within and among natural populations. University of Chicago press.
41. Hartl DL, Clark AG, Clark AG. 1997. Principles of population genetics. Sinauer associates Sunderland, MA.
42. Sun Y, Luo H. 2018. Homologous recombination in core genomes facilitates marine bacterial adaptation. *Appl Environ Microbiol* 84:e02545-17. <https://doi.org/10.1128/AEM.02545-17>.
43. Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, Chen F, Lapidus A, Ferreira S, Johnson J, Steglich C, Church GM, Richardson P, Chisholm SW. 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* 3:e231. <https://doi.org/10.1371/journal.pgen.0030231>.
44. Giovannoni SJ, Thrash JC, Temperton B. 2014. Implications of streamlining theory for microbial ecology. *ISME J* 8:1553–1565. <https://doi.org/10.1038/ismej.2014.60>.
45. 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>.
46. Abou-El-Wafa GSE, Shaaban M, Shaaban KA, El-Naggar MEE, Maier A, Fiebig HH, Laatsch H. 2013. Pachydictyols B and C: new diterpenes from *Dictyota dichotoma* Hudson. *Mar Drugs* 11:3109–3123. <https://doi.org/10.3390/md11093109>.
47. Bourgougnon N, Stiger-Pouvreau V. 2011. Chemodiversity and bioactivity within red and brown macroalgae along the French coasts, metropole and overseas departments and territories. In *Handbook of Marine Macroalgae* 2011. 58–105. John Wiley & Sons, Ltd.
48. Zhuang G-C, Peña-Montenegro TD, Montgomery A, Montoya JP, Joye SB. 2019. Significance of acetate as a microbial carbon and energy source in the water column of Gulf of Mexico: implications for marine carbon cycling. *Global Biogeochem Cycles* 33:223–235. <https://doi.org/10.1029/2018GB006129>.
49. Ghadriyanfar M, Rosentrater KA, Keyhani A, Omid M. 2016. A review of macroalgae production, with potential applications in biofuels and bioremediation. *Renew Sust Energ Rev* 54:473–481. <https://doi.org/10.1016/j.rser.2015.10.022>.
50. De Jesus Raposo MF, De Morais AMB, De Morais RMSC. 2015. Marine polysaccharides from algae with potential biomedical applications. *Mar Drugs* 13:2967–3028. <https://doi.org/10.3390/md13052967>.
51. Hamid SS, Wakayama M, Ichihara K, Sakurai K, Ashino Y, Kadowaki R, Soga T, Tomita M. 2019. Metabolome profiling of various seaweed species discriminates between brown, red, and green algae. *Planta* 249:1921–1947. <https://doi.org/10.1007/s00425-019-03134-1>.
52. Lim SJ, Aida WMW, Schiehser S, Rosenau T, Böhmendorfer S. 2019. Structural elucidation of fucoidan from *Cladosiphon okamuranus* (Okinawa mozuku). *Food Chem* 272:222–226. <https://doi.org/10.1016/j.foodchem.2018.08.034>.
53. Hosler JP, Fetter J, Tecklenburg MM, Espe M, Lerma C, Ferguson-Miller S. 1992. Cytochrome aa₃ of *Rhodobacter sphaeroides* as a model for mitochondrial cytochrome c oxidase. Purification, kinetics, proton pumping, and spectral analysis. *J Biol Chem* 267:24264–24272. [https://doi.org/10.1016/S0021-9258\(18\)35760-0](https://doi.org/10.1016/S0021-9258(18)35760-0).
54. Pitcher RS, Watmough NJ. 2004. The bacterial cytochrome cbb₃ oxidases. *Biochim Biophys Acta* 1655:388–399. <https://doi.org/10.1016/j.bbabbio.2003.09.017>.
55. Pitcher RS, Brittain T, Watmough NJ. 2002. Cytochrome cbb₃ oxidase and bacterial microaerobic metabolism. *Biochem Soc Trans* 30:653–658. <https://doi.org/10.1042/bst0300653>.
56. Li Y, Raschdorf O, Silva KT, Schüler D. 2014. The terminal oxidase cbb₃ functions in redox control of magnetite biomineralization in *Magnetospirillum gryphiswaldense*. *J Bacteriol* 196:2552–2562. <https://doi.org/10.1128/JB.101652-14>.
57. Ekici S, Pawlik G, Lohmeyer E, Koch H-G, Daldal F. 2012. Biogenesis of cbb₃-type cytochrome c oxidase in *Rhodobacter capsulatus*. *Biochim Biophys Acta* 1817:898–910. <https://doi.org/10.1016/j.bbabbio.2011.10.011>.
58. Koch H-G, Hwang O, Daldal F. 1998. Isolation and characterization of *Rhodobacter capsulatus* mutants affected in cytochrome cbb₃ oxidase activity. *J Bacteriol* 180:969–978. <https://doi.org/10.1128/JB.180.4.969-978.1998>.
59. Moran MA, Buchan A, González JM, Heidelberg JF, Whitman WB, Kiene RP, Henriksen JR, King GM, Belas R, Fuqua C, Brinkac L, Lewis M, Johri S, Weaver B, Pai G, Eisen JA, Rahe E, Sheldon WM, Ye W, Miller TR, Carlton J, Rasko DA, Paulsen IT, Ren Q, Daugherty SC, Deboy RT, Dodson RJ, Durkin AS, Madupu R, Nelson WC, Sullivan SA, Rosovitz MJ, Haft DH, Selengut J, Ward N. 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature* 432:910–913. <https://doi.org/10.1038/nature03170>.
60. Button DK. 1998. Nutrient uptake by microorganisms according to kinetic parameters from theory as related to cytoarchitecture. *Microbiol Mol Biol Rev* 62:636–645. <https://doi.org/10.1128/MMBR.62.3.636-645.1998>.
61. Button DK. 1991. Biochemical basis for whole-cell uptake kinetics: specific affinity, oligotrophic capacity, and the meaning of the Michaelis constant. *Appl Environ Microbiol* 57:2033–2038. <https://doi.org/10.1128/aem.57.7.2033-2038.1991>.
62. Haas AF, Fairoz MFM, Kelly LW, Nelson CE, Dinsdale EA, Edwards RA, Giles S, Hatay M, Hisakawa N, Knowles B, Lim YW, Maughan H, Pantos O, Roach TNF, Sanchez SE, Silveira CB, Sandin S, Smith JE, Rohwer F. 2016. Global microbialization of coral reefs. *Nat Microbiol* 1:1–7. <https://doi.org/10.1038/nmicrobiol.2016.42>.
63. Barott KL, Rohwer FL. 2012. Unseen players shape benthic competition on coral reefs. *Trends Microbiol* 20:621–628. <https://doi.org/10.1016/j.tim.2012.08.004>.
64. Chen J, Li H, Zhang Z, He C, Shi Q, Jiao N, Zhang Y. 2020. DOC dynamics and bacterial community succession during long-term degradation of *Ulva prolifera* and their implications for the legacy effect of green tides on refractory DOC pool in seawater. *Water Res* 185:116268. <https://doi.org/10.1016/j.watres.2020.116268>.
65. Silveira CB, Luque A, Roach TN, Villela H, Barno A, Green K, Reyes B, Rubio-Portillo E, Le T, Mead S, Hatay M, Vermeij MJ, Takeshita Y, Haas A, Bailey B, Rohwer F. 2019. Biophysical and physiological processes causing oxygen loss from coral reefs. *Elife* 8:e49114. <https://doi.org/10.7554/eLife.49114>.
66. Cabanillas-Terán N, Hernández-Arana HA, Ruiz-Zárata M-Á, Vega-Zepeda A, Sanchez-Gonzalez A. 2019. Sargassum blooms in the Caribbean alter the trophic structure of the sea urchin *Diadema antillarum*. *PeerJ* 7:e7589. <https://doi.org/10.7717/peerj.7589>.
67. Luo H, Moran MA. 2015. How do divergent ecological strategies emerge among marine bacterioplankton lineages? *Trends Microbiol* 23:577–584. <https://doi.org/10.1016/j.tim.2015.05.004>.
68. Dang H, Lovell CR. 2016. Microbial surface colonization and biofilm development in marine environments. *Microbiol Mol Biol Rev* 80:91–138. <https://doi.org/10.1128/MMBR.00037-15>.
69. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
70. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from

- isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
71. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
 72. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The rast server: rapid annotations using subsystems technology. *BMC Genomics* 9:1–15. <https://doi.org/10.1186/1471-2164-9-75>.
 73. Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L, He S, Hurwitz DI, Jackson JD, Ke Z, Krylov D, Lanczycki CJ, Liebert CA, Liu C, Lu F, Lu S, Marchler GH, Mullokandov M, Song JS, Thanki N, Yamashita RA, Yin JJ, Zhang D, Bryant SH. 2007. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* 35:D237–D240. <https://doi.org/10.1093/nar/gkl951>.
 74. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/nar/gkv1070>.
 75. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32:D277–D280. <https://doi.org/10.1093/nar/gkh063>.
 76. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. <https://doi.org/10.1186/1471-2105-4-41>.
 77. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, Brinkman FS, Group SFURC. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* 45:W30–W35. <https://doi.org/10.1093/nar/gkx343>.
 78. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the phast phage search tool. *Nucleic Acids Res* 44:W16–W21. <https://doi.org/10.1093/nar/gkv387>.
 79. Siguier P, Pérochon J, Lestrade L, Mahillon J, Chandler M. 2006. Isfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34:D32–D36. <https://doi.org/10.1093/nar/gkj014>.
 80. Lerat E, Ochman H. 2004. Ψ - Φ : exploring the outer limits of bacterial pseudogenes. *Genome Res* 14:2273–2278. <https://doi.org/10.1101/gr.2925604>.
 81. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
 82. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43:e15. <https://doi.org/10.1093/nar/gku1196>.
 83. Arevalo P, Vanlnsberghe D, Elsherbini J, Gore J, Polz MF. 2019. A reverse ecology approach based on a biological definition of microbial populations. *Cell* 178:820–834. <https://doi.org/10.1016/j.cell.2019.06.033>.
 84. Bobay L-M, Ellis BS-H, Ochman H. 2018. Conspicifex: classifying prokaryotic species based on gene flow. *Bioinformatics* 34:3738–3740. <https://doi.org/10.1093/bioinformatics/bty400>.
 85. Excoffier L, Lischer HE. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under linux and windows. *Mol Ecol Resour* 10:564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>.
 86. Kolde R, Kolde MR. 2015. Package ‘pheatmap.’ R package 1:790.
 87. Bochner BR, Gadzinski P, Panomitros E. 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome Res* 11:1246–1255. <https://doi.org/10.1101/gr.186501>.
 88. Gong X, Garcia-Robledo E, Schramm A, Revsbech NP. 2015. Respiratory kinetics of marine bacteria exposed to decreasing oxygen concentrations. *Appl Environ Microbiol* 82:1412–1422. <https://doi.org/10.1128/AEM.03669-15>.
 89. Garcia-Jimenez P, Llorens C, Roig FJ, Robaina RR. 2018. Analysis of the transcriptome of the red seaweed *Grateloupia imbricata* with emphasis on reproductive potential. *Mar Drugs* 16(12):490.
 90. He Y, Hu C, Wang Y, Cui D, Sun X, Li Y, Xu N. 2018. The metabolic survival strategy of marine macroalga *Ulva prolifera* under temperature stress. *J Appl Phycol* 30:3611–3621.