



# Homologous Recombination in Core Genomes Facilitates Marine Bacterial Adaptation

Ying Sun,<sup>a</sup> Haiwei Luo<sup>a</sup>

<sup>a</sup>Simon F. S. Li Marine Science Laboratory of School of Life Sciences and Partner State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong

**ABSTRACT** Acquisition of ecologically relevant genes is common among ocean bacteria, but whether it has a major impact on genome evolution in marine environments remains unknown. Here, we analyzed the core genomes of 16 phylogenetically diverse and ecologically relevant bacterioplankton lineages, each consisting of up to five genomes varying at the strain level. Statistical approaches identified from each lineage up to ~50 loci showing anomalously high divergence at synonymous sites, which is best explained by recombination with distantly related organisms. The enriched gene categories in these outlier loci match well with the characteristics previously identified as the key phenotypes of these lineages. Examples are antibiotic synthesis and detoxification in *Phaeobacter inhibens*, exopolysaccharide production in *Alteromonas macleodii*, hydrocarbon degradation in *Marinobacter hydrocarbonoclasticus*, and cold adaptation in *Pseudoalteromonas haloplanktis*. Intriguingly, the outlier loci feature polysaccharide catabolism in *Cellulophaga baltica* but not in *Cellulophaga lytica*, consistent with their primary habitat preferences in macroalgae and beach sands, respectively. Likewise, analysis of *Prochlorococcus* showed that photosynthesis-related genes listed in the outlier loci are found only in the high-light-adapted ecotype and not in the low-light adapted ecotype. These observations strongly suggest that recombination with distant relatives is a key mechanism driving the ecological diversification among marine bacterial lineages.

**IMPORTANCE** Acquisition of new metabolic genes has been known as an important mechanism driving bacterial evolution and adaptation in the ocean, but acquisition of novel alleles of existing genes and its potential ecological role have not been examined. Guided by population genetic theories, our genomic analysis showed that divergent allele acquisition is prevalent in phylogenetically diverse marine bacterial lineages and that the affected loci often encode metabolic functions that underlie the known ecological roles of the lineages under study.

**KEYWORDS** homologous recombination, lateral gene transfer, *Roseobacter*, SAR11, *Flavobacteria*, *Prochlorococcus*, *Synechococcus*, *Pseudoalteromonas*, *Marinobacter*, *Alteromonas*, *Glaciecola*

Lateral gene transfer (LGT) has been repeatedly proposed as a major mechanism giving rise to marine microbially diverse populations. A high frequency of LGT occurs in marine bacterioplankton lineages through conjugation (1), transduction (2, 3), genetic transformation (4), and gene transfer agents (5). More recently, the most abundant free-living marine cyanobacteria, *Prochlorococcus* spp., were found to produce abundant extracellular membrane vesicles (6), which are known as important media for LGT in *Escherichia coli* (7) and *Acinetobacter* (8). LGT may transfer a novel gene that did not occur in the recipient genome. It may also replace the original copy with a novel allele at the same locus, a mechanism called homologous recombination (HR). Regardless of these possibilities, LGT is a rapid way to confer new functions to the recipient organisms and may enable new biogeochemical transformations, but it is not

**Received** 15 November 2017 **Accepted** 20 March 2018

**Accepted manuscript posted online** 23 March 2018

**Citation** 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>.

**Editor** Haruyuki Atomi, Kyoto University

**Copyright** © 2018 American Society for Microbiology. All Rights Reserved.

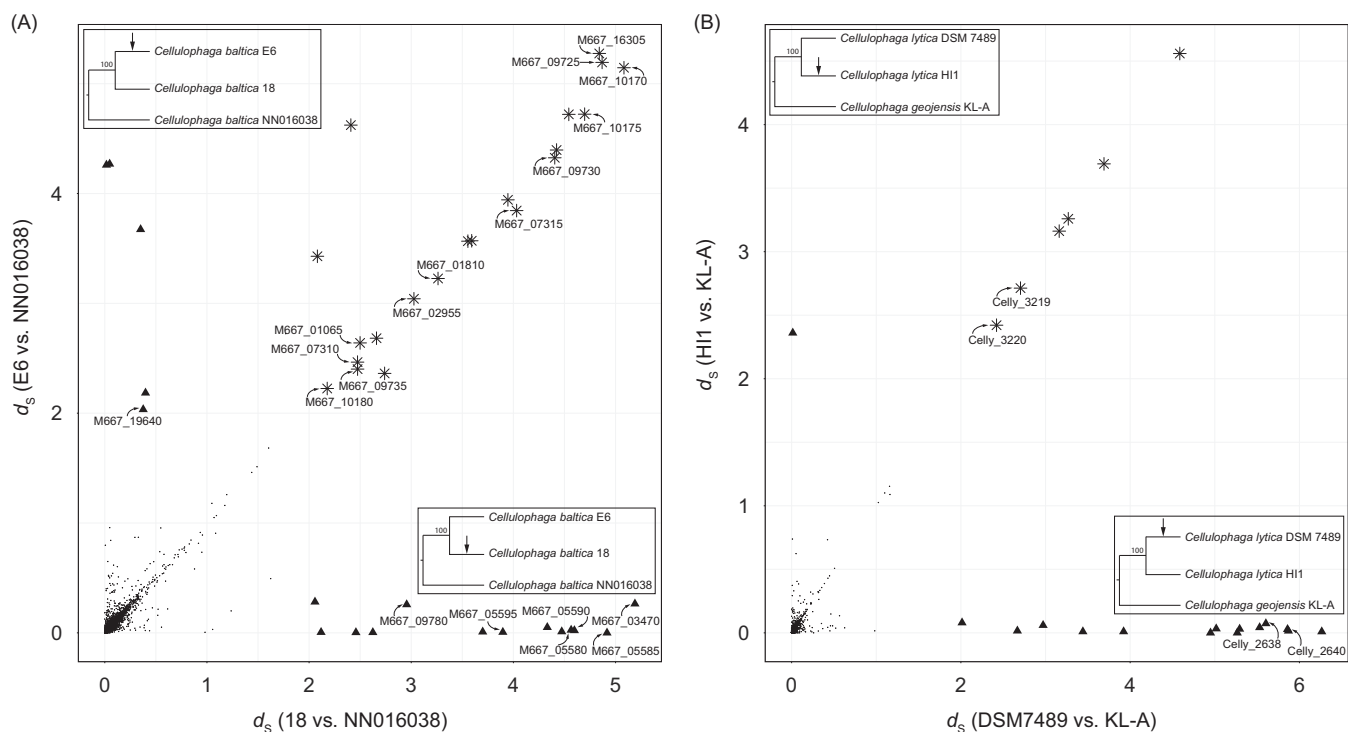
Address correspondence to Haiwei Luo, [hluo2006@gmail.com](mailto:hluo2006@gmail.com).

clear whether incidences of LGT events have a major impact on genome evolution and diversification of bacterial lineages in the ocean.

To answer these questions, we explored the recombination pattern occurring in the core genomic regions of marine bacterial lineages, each varying at the strain level. Here, the core genome consists of single-copy genes shared by all members of a lineage. Because the core genomic regions represent 82.01% ( $\pm 7.23\%$ ) of the whole genome among the 16 examined lineages in *Alphaproteobacteria*, *Gammaproteobacteria*, *Flavobacteria*, and *Cyanobacteria* (see Table S1 in the supplemental material), any major recombination pattern in the core genome may have consequences in microbial evolution at the genomic scale. Since synonymous mutations of a protein-coding gene do not lead to amino acid replacement, they are often considered neutral or nearly so. Without recombination, divergence at synonymous sites between two orthologous genes is determined largely by the mutation rate and their divergence time, and most of the variation of synonymous divergence among loci can be explained by the stochastic nature of mutations among genomic regions. If a divergent allele was introduced through recombination, which is not uncommon in prokaryotic lineages, the synonymous site of the recombined loci would expectedly show an unusually large synonymous substitution rate ( $d_s$ ) (defined as the number of synonymous substitutions per synonymous site) compared to those of the remaining regions of the genome (9). Compared to nonsynonymous changes, which are subject to strong functional constraints at the protein sequence level, synonymous changes are largely neutral in a vast majority of genes (10, 11), and thus enormously large  $d_s$  values provide a reliable measure of allele acquisition from divergent relatives (12–15). It is worth mentioning that the assumption that changes at synonymous sites are effectively neutral has been questioned. Our recent studies have shown that nutrient limitations act as selective pressures driving genomic G+C content evolution in marine bacterioplankton lineages (16, 17). Under these mechanisms, synonymous sites are affected indiscriminately among all genes, and thus unusual  $d_s$  values are unlikely to arise in a small subset of gene families. In fact, identifying unusual  $d_s$  values as a way to find genes subject to recombination has been used in several studies (12, 15, 18). Here, we identified such loci and linked their predicted functions to the ecological roles of these lineages.

## RESULTS AND DISCUSSION

**Recombination with distant relatives in marine *Flavobacteria* lineages.** As described in Materials and Methods, potential allele acquisition events from distant relatives were differentiated according to the phylogenetic branches where recombination may have occurred. In the case of *Cellulophaga baltica*, 21 gene families were potentially subject to recombination occurring at the outgroup branch (NN016038) (shown as asterisks in Fig. 1A; see Fig. S1A and Table S2A in the supplemental material). These loci are highly enriched with polysaccharide utilization genes (Fig. 1A; Table S2A). This includes a glycoside hydrolase (GH) cluster (M667\_09725, \_09730, and \_09735) consisting of genes encoding  $\alpha$ -1,3-(3,6-anhydro)-L-galactosidase, polygalacturonase, and  $\beta$ -porphyranase, which hydrolyze the glycosidic bonds in the linear polysaccharide chains of agarose, pectin, and porphyran, respectively (19). Phylogenetic analyses showed that these three neighboring GH genes each have a distinct evolutionary history. For instance, *Formosa agariphila* KMM 3901, a close relative of *Cellulophaga baltica*, is more closely related to *C. baltica* 18 and *C. baltica* E6 for  $\alpha$ -1,3-(3,6-anhydro)-L-galactosidase (see Fig. S2A in the supplemental material) but instead becomes more closely related to *C. baltica* NN016038 for polygalacturonase (Fig. S2B) and  $\beta$ -porphyranase (Fig. S2C), suggesting that recombination occurred independently among these GH genes. Furthermore, this GH cluster is flanked by genes involved in the degradation of various polysaccharides, including multiple copies of genes for sulfatase, peptidylglycine monooxygenase,  $\beta$ -agarase, and alpha-L-fucosidase, as well as tandem *susCD*-like genes, which together form a large polysaccharide utilization locus (PUL) (Fig. S2D).

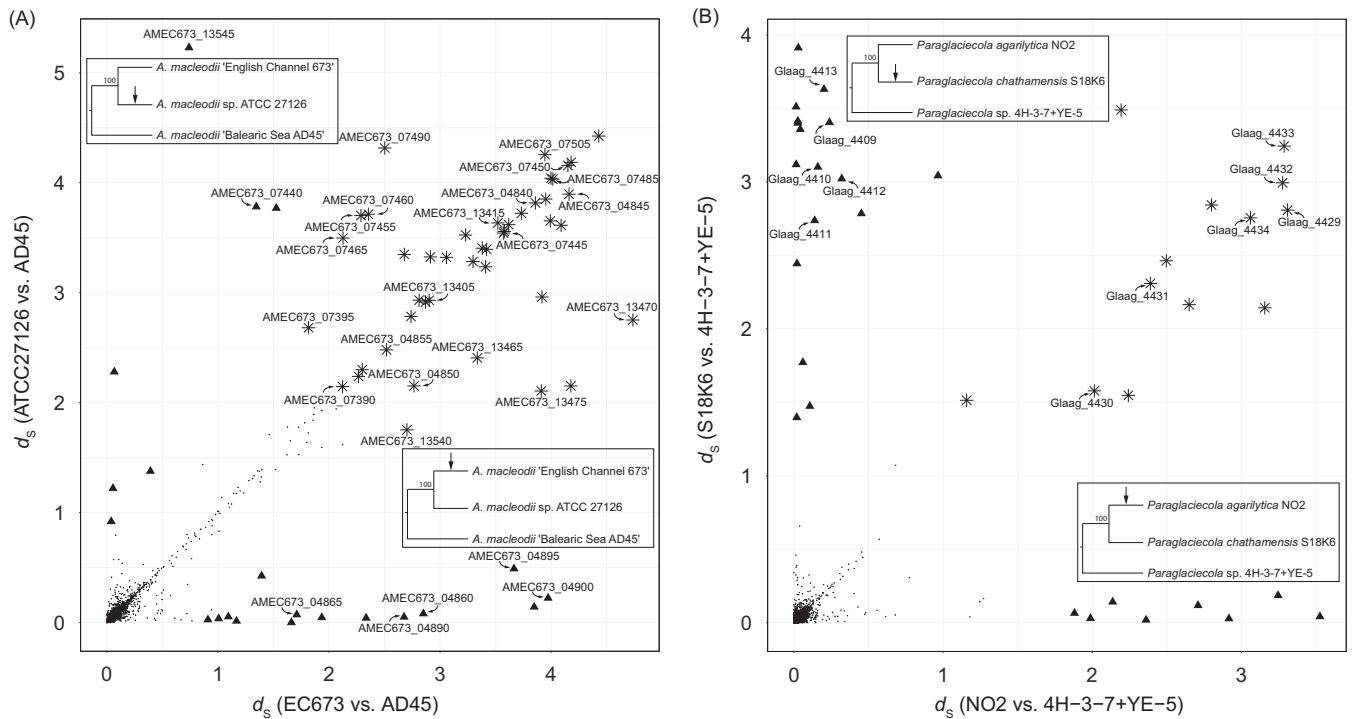


**FIG 1** Detection of the  $d_s$  outliers of *Cellulophaga baltica* (A) and *Cellulophaga lytica* (B) through two rounds of linear regression analysis. Here,  $d_s$  was calculated between each of the ingroup genomes and the outgroup genome across all single-copy orthologous genes shared by the three genomes. Take *Cellulophaga baltica* as an example. Strain NN016038 is the outgroup, and strains E6 and "18" comprise the ingroup. The  $d_s$  was calculated between NN016038 and E6 and between NN016038 and "18." The first linear regression (explained in the legend to Fig. S1 in the supplemental material) led to the identification of the outlier genes in which divergent alleles were acquired at the outgroup branch. These outlier genes are shown in Fig. S1A and marked with asterisks in panel A. The second linear regression was performed using the  $d_s$  of E6/NN016038 against the  $d_s$  of "18"/NN016038 after the genes shown with asterisks were excluded. It led to the identification of outlier genes resulting from divergent allele acquisitions at one of the ingroup branches, and these outliers are marked as solid triangles. Schematic representations of the two possible allele acquisition scenarios at ingroup branches are also shown: an acquisition at E6 (top left) and "18" (bottom right) in panel A and an acquisition at HI1 (top left) and DSM 7489 (bottom right) in panel B. The locus tags for ecologically relevant outlier genes are shown next to the outlier gene symbols, and their predicted functions are given in Table S2 in the supplemental material.

The loci subject to recombination at the outgroup branch also highlight genes encoding iota-carrageenase (M667\_16305), which degrades the red algal cell wall polysaccharide iota-carrageenan, a glycan metabolism protein (M667\_07315), and four SusC-like TonB-dependent transporters (M667\_01065, \_01810, \_02955, and \_07310) that are known members in the PUL-like systems and function in concert with linked GHs (20, 21). In addition, many more outlier genes are involved in sugar metabolism. These include a gene cluster for sugar transport (M667\_10170, \_10175, and \_10180) and several loci showing evidence of recombination with distant relatives occurring at the ingroup branches (shown as triangles in Fig. 1A), such as a gene cluster for sugar synthesis (M667\_05580, \_05585, \_05590, and \_05595), a sugar transferase (M667\_03470), a glycosyl hydrolase (M667\_09780), and a carbohydrate-binding protein (M667\_19640) (Table S2A).

In the case of the closely related lineage *Cellulophaga lytica* (Fig. 1B; Fig. S1B), polysaccharide-catabolizing genes are surprisingly depleted in the gene pools showing an unusual pattern of  $d_s$  (Table S2B). Instead, two genes engaged in cobalt/zinc/cadmium resistance (Celly\_3219 and \_3220) and a few genes involved in the biosynthesis of the O-antigen precursors UDP- $\alpha$ -D-glucuronate (Celly\_2638) and UDP- $\alpha$ -D-galactose (Celly\_2640) show evidence for recombination with distant relatives.

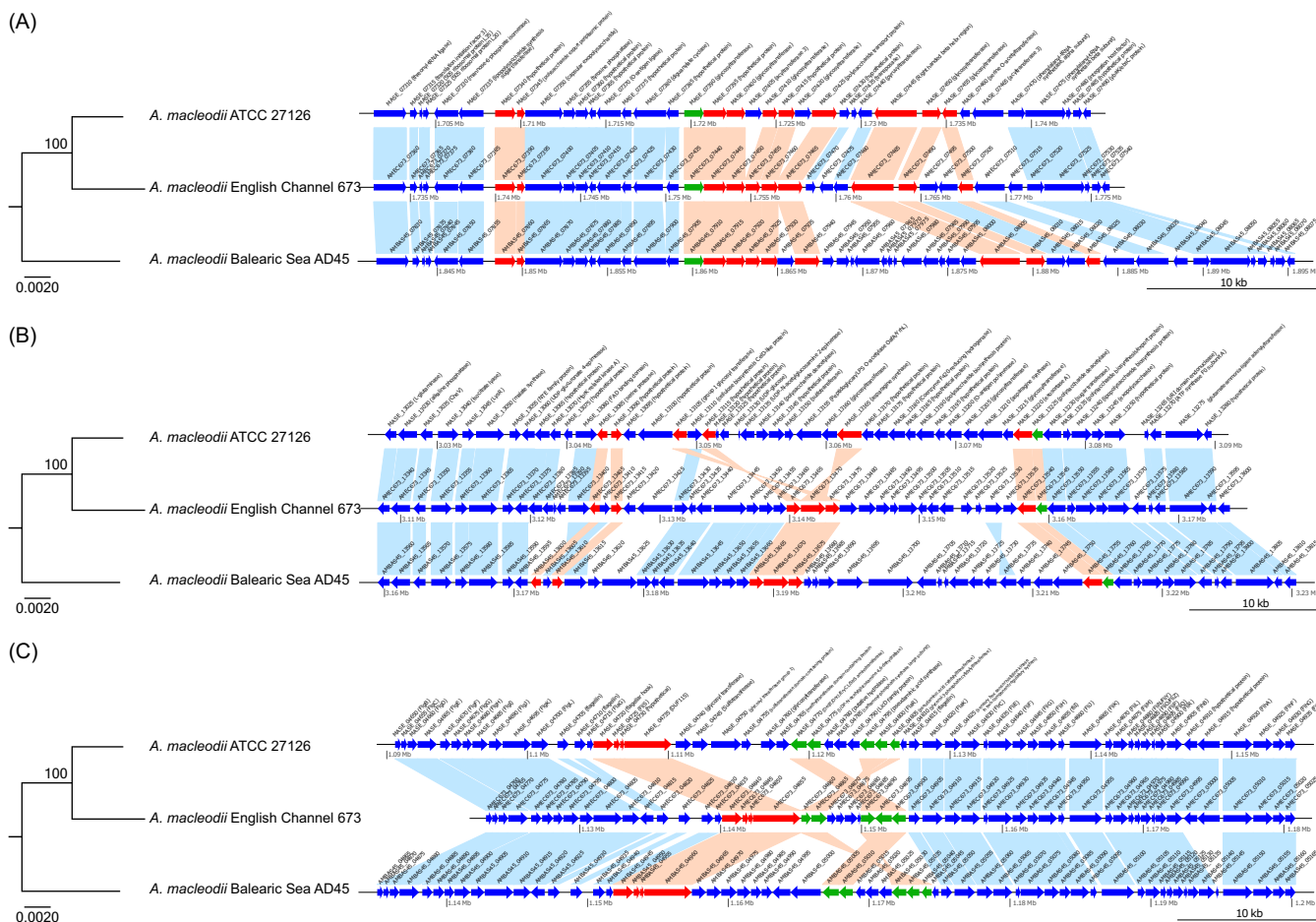
**Recombination with distant relatives in marine *Alteromonadales* lineages.** Application of these approaches to the three closely related genomes of *Alteromonas macleodii* led to the identification of 44 gene families showing evidence of potential recombination with distant relatives at the outgroup branch and 21 families of recombination at the ingroup branches (Fig. 2A; see Fig. S1C and Table S3A in the supple-



**FIG 2** Detection of  $d_s$  outliers of *Alteromonas macleodii* (A) and *Paraglaucicocla agarilytica* (B) through the linear regression procedure outlined in the Fig. 1 legend. Asterisks and solid triangles represent outlier genes in which the divergent alleles were acquired at the outgroup and ingroup branches, respectively. The predicted functions of outlier genes are given in Table S3 in the supplemental material.

mental material). One major characteristic is that these genes are often clustered and located in previously reported genomic islands (GIs) (Table S3A) involved in cell surface modifications mainly for exopolysaccharide (EPS) production and for phage avoidance (22, 23). The  $d_s$  outlier genes related to capsular EPS biosynthesis include genes for a polysaccharide export outer membrane protein (AMEC673\_07395), five glycosyltransferases (AMEC673\_07440, \_07450, \_07455, \_07490, and \_07505), and a putative polysaccharide transporter (AMEC673\_07465) (Fig. 3A). Capsular polysaccharide and exopolysaccharide are recognized media for cell-cell communication; they are critical compounds involved in adhesion for surface colonization, cohesion of biofilm, tolerance to antimicrobial agents, resistance to phage infection, and defense against protozoan attack, among other functions (24). Our finding that the  $d_s$  outlier list includes genes involved in EPS metabolism is consistent with observations that *A. macleodii* is among the most potent bacteria for EPS production (25), and it is commonly used to produce EPS for industry and other purposes (25, 26).

One prominent example related to phage avoidance in the  $d_s$  outlier list is the cluster involved in the synthesis of the O chain of lipopolysaccharide (LPS) (AMEC673\_13405, \_13415, \_13465, \_13470, \_13475, \_13540, and \_13545) within the LPS O-chain GI proposed previously (23) (Fig. 3B). The LPS O chain has been known as a phage receptor (27), and recombination with divergent relatives is likely an adaptive strategy to escape from phage recognition (28). In fact, high diversity of these loci has been postulated to be an important mechanism for phage avoidance in *A. macleodii* (23). Another gene cluster, located inside a flagellum glycosylation island, encodes the major exposed components of the flagellum, including two flagellar proteins (AMEC673\_04840 and \_04845), and an additional gene cluster includes genes for the *N*-acetylneuraminyl transferase (AMEC673\_04860) essential for the formation of the capsule, the UDP-2,4-diacetamido-2,4,6-trideoxy-beta-L-altropyranose hydrolase (AMEC673\_04865) and the pseudaminic acid synthase (AMEC673\_04890) used for flagellin modification (29), the DegT/DnrJ/EryC1/StrS aminotransferase (AMEC673\_



**FIG 3** Three loci with anomalously high  $d_s$  values located in previously reported genomic islands (GIs) of exopolysaccharide (EPS) biosynthesis (A), lipopolysaccharide (LPS) O-chain production (B), and flagellum glycosylation (C) in *Alteromonas macleodii*. Loci with unusually high  $d_s$  values are indicated by red (recombination at the outgroup branch) and green (recombination at an ingroup branch) arrows, and the remaining genes are indicated by blue arrows. Orthologous genes between genomes are linked with shades of blue for genes with anomalously high  $d_s$  values and blue for the rest. Parallelograms illustrate relocations on the same strand, while a twist represents an inversion. Phylograms on the left of each panel describe the relationships among the three taxa.

04895), and the nucleoside-diphosphate sugar epimerases (AMEC673\_04900) engaged in polysaccharide biogenesis (Fig. 3C). The flagellum has been suggested to be the target for recognition by phages in *A. macleodii* (23), and flagellum-specific phages may interact with their host by attachment to flagella followed by DNA injection (30). Our analysis suggests that recombination with divergent alleles may allow *A. macleodii* to escape from phage infection.

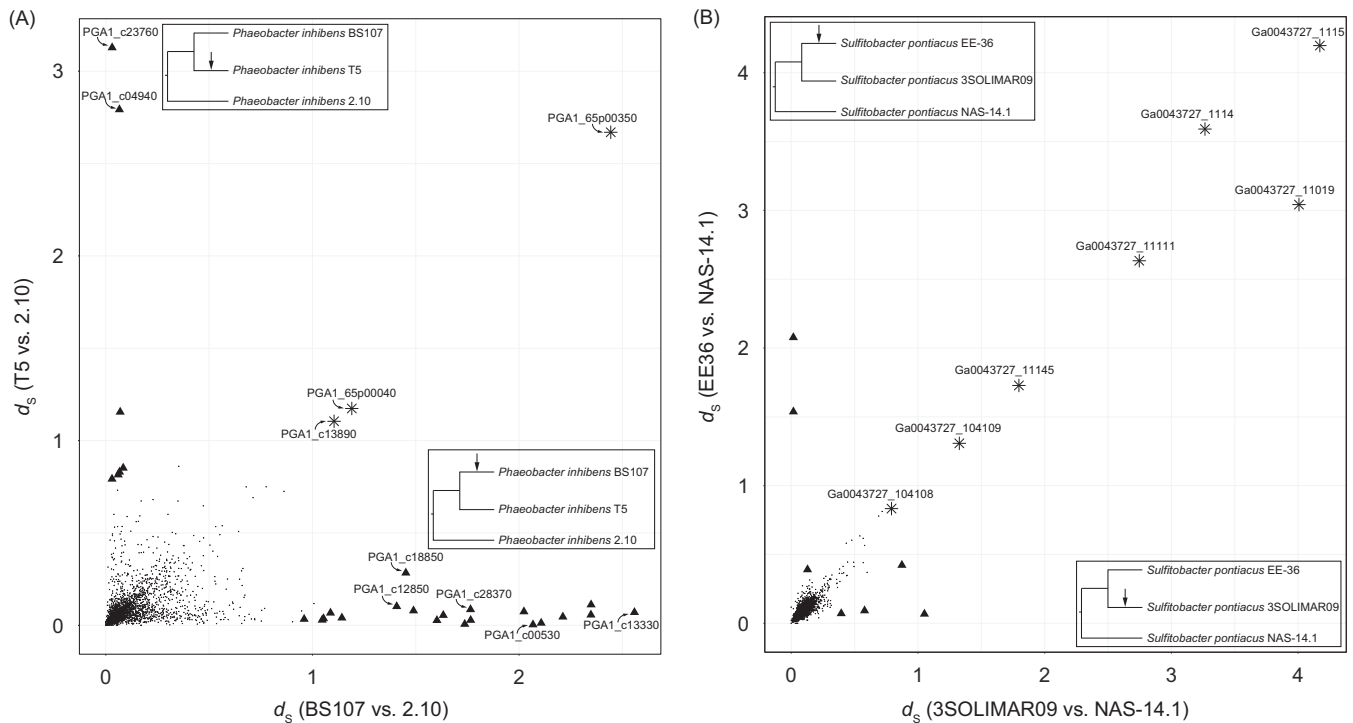
In contrast, the  $d_s$  outlier genes in a related lineage, *Paraglaucicola chathamensis* (Fig. 2B; Fig. S1D), feature heavy metal resistance functions (Table S3B). This includes a gene cluster (Glaag\_4429, \_4430, \_4431, \_4432, \_4433, \_4434, and \_4435) responsible for resistance to cobalt, zinc, and cadmium. Similar *czc* cassettes for metal efflux pumps are broadly present in *Alteromonas* genomes and may be used to export toxic products that are imported along with nutrient uptake from particulate organic matter (23). It also includes a gene cluster responsible for mercury (Hg) resistance, consisting of genes for an Hg-responsive transcriptional regulator (Glaag\_4413), a periplasmic mercuric ion-scavenging protein (Glaag\_4411), two inner membrane-spanning Hg(II) and phenylmercury transporters (Glaag\_4410 and \_4412), and an Hg(II) reductase (Glaag\_4409) (Table S3B). Further analyses suggest that recombination at these Hg resistance loci occurred on the leaf branch of *Paraglaucicola chathamensis* S18K6 (see Fig. S3 in the supplemental material). This strain was isolated from the deep-sea sediment near

Chatham Rise (31), a region where numerous seamounts with hydrothermal vents are distributed (32). One important feature of the vent habitats is high levels of heavy metals, including Hg, and indeed an Hg-resistant bacterium was isolated from a similar vent environment (33). Our finding therefore suggests that allele acquisition from distant relatives at loci involved in Hg metabolism is potentially an important adaptive mechanism for *P. chathamensis* S18K6.

A third related lineage, *Marinobacter hydrocarbonoclasticus*, contains 56 gene families showing unusual  $d_s$  values (see Fig. S4A and Tables S3C and S4A in the supplemental material). Among them, genes involved in aromatic hydrocarbon degradation (MARHY0454) and flagellar assembly (MARHY2506, -2508, and -2178) were subject to repeated recombination with divergent alleles (see Fig. S5A, panels i, g, and h, in the supplemental material). These outlier genes also feature clustered loci responsible for signal transduction, including a cluster (MARHY3575, -3576, -3577, and -3578) encoding the chemotaxis protein CheY, diguanylate cyclase, a two-component system sensor histidine kinase, and another histidine kinase and a second cluster (MARHY1182, -1183, and -1184) encoding a chemotaxis protein, a pseudouridylate synthase, and a diguanylate cyclase. In addition, they include genes for the enzyme glutathione *S*-transferase (MARHY2825), which has been used as a biomarker of polycyclic aromatic hydrocarbon (PAH) pollution (34), a cluster (MARHY3259, -3260, -3261, and -3262) involved in the catabolism of hydrophobic carbons, including an alkylhydroperoxidase, an alcohol dehydrogenase, a TetR family transcriptional regulator, and an aldehyde dehydrogenase, a cluster (MARHY0073 and -0075) for a dicarboxylic acid transport system, and a cluster (MARHY0216, -0217, -0218, and -0219) encoding a TonB-dependent receptor and ABC transporters. The four strains used here were all isolated from oil-contaminated beach sands (35, 36) and are capable of degrading a wide range of hydrophobic organic compounds (37). Previous metatranscriptomic studies showed that genes involved in PAH, alkane, and toluene degradation, as well as motility and chemotaxis, are often highly expressed in natural beach communities associated with oil contamination (38, 39). Because the genes identified through the  $d_s$  outlier approaches agree well with the highly expressed genes, we suggest that recombination with distant relatives is potentially an important mechanism increasing their capabilities in hydrocarbon degradation.

A fourth lineage, *Pseudoalteromonas haloplanktis*, contains 67 gene families showing an anomalous  $d_s$  pattern (Fig. S4B; Tables S3D and S4B). The most interesting gene within this outlier list is PH505\_be00240, an ortholog of the gene for a cold-adapted *S*-formylglutathione hydrolase (PSHAa1385) in the closely related *P. haloplanktis* TAC125 (see Fig. S6 in the supplemental material). The crystal structure of this protein from *P. haloplanktis* TAC125 revealed a significant reduction of arginine residues and iron pairs, as well as a lower arginine/(arginine + lysine) ratio, providing appropriate flexibility around the active site for catalysis at low temperatures (40). Other interesting outlier genes related to cold adaptation include a gene (PH505\_ap00150) orthologous to a cold shock RNA methyltransferase gene (PSHAb0516) in *P. haloplanktis* TAC125 (41) and three genes involved in the defense against oxidative stress, including genes for a peroxiredoxin (PH505\_dz00090) responsible for the thiol-dependent reduction of peroxides (42), a chorismate pyruvate-lyase (PH505\_bj00180) engaged in the biosynthesis of ubiquinone, which is a lipid subject to oxido-reduction cycles (43), and a lactoylglutathione lyase (PH505\_dn00020) transforming methylglyoxal and glutathione to *S*-lactoylglutathione in methylglyoxal degradation (44). These genes are potentially important for *P. haloplanktis* to adopt a psychrophilic lifestyle, suggesting that acquisition of divergent alleles is an adaptive mechanism for cold adaptation.

**Recombination with distant relatives in marine *Roseobacter* lineages.** Using the same  $d_s$  outlier approaches, we investigated two marine *Roseobacter* species, each consisting of three genomes. In the case of *Phaeobacter inhibens*, three gene families were potentially subject to recombination occurring at the outgroup branch (PG210) (shown as asterisks in Fig. 4A; see Fig. S1E and Table S5A in the supplemental material).



**FIG 4** Detection of  $d_s$  outliers of *Phaeobacter inhibens* (A) and *Sulfotobacter pontiacus* (B) through the linear regression procedure outlined in the Fig. 1 legend. Asterisks and solid triangles represent outlier genes in which the divergent alleles were acquired at the outgroup and ingroup branches, respectively. The predicted functions of outlier genes are given in Table S5 in the supplemental material.

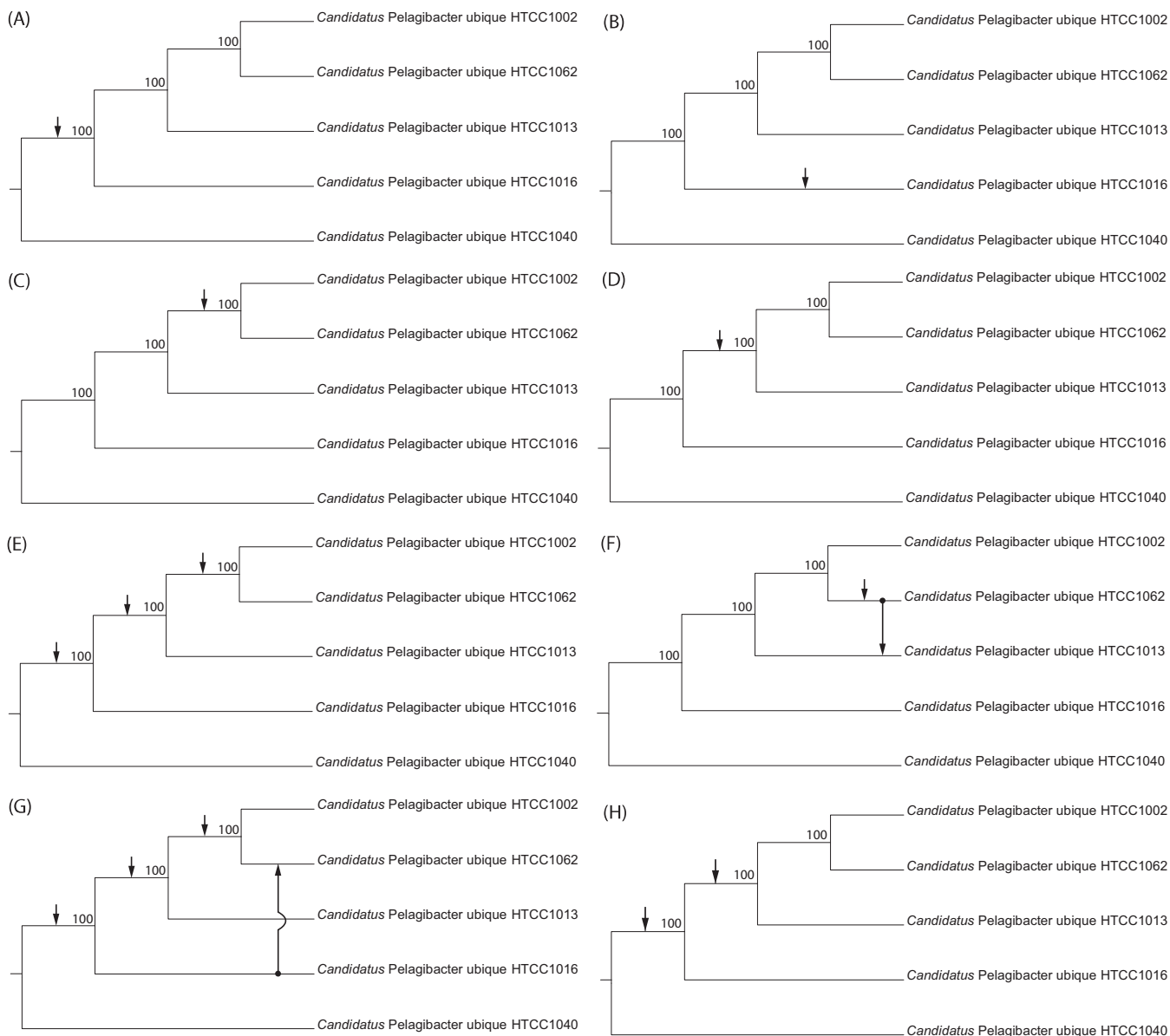
These include two genes encoding repeats-in-toxin (RTX) proteins (PGA1\_65p00040 and \_65p00350) that function as virulence factors (45) and a third gene encoding lactoylglutathione lyase (PGA1\_c13890), which is responsible for detoxification of methylglyoxal, a highly toxic by-product of glycolysis metabolism (46) contributing to intracellular invasiveness (47). In addition, 27 gene families show evidence of potential recombination with distant relatives occurring at the ingroup branches (shown as triangles in Fig. 4A). These outlier genes were highly enriched in antibiotic production and resistance (Table S5A). One example is aminoglycoside phosphotransferase (PGA1\_c00530), which catalyzes the phosphorylation of the aminoglycoside and macrolide antibiotics, leading to their inactivation and to bacterial antibiotic resistance (48). A second example is D-hydantoinase (PGA1\_c12850), which catalyzes the hydrolysis of 5-substituted hydantoin in microorganisms, leading to enantiomerically pure *N*-carbamyl amino acids, and is widely used for the production of antibiotics, peptide hormones, pyrethroids, and pesticides (49). A third example is the deacylase PagL (PGA1\_c04940), which is important for virulence and resistance to cationic antimicrobial peptides via regulating lipid A modification (50). A fourth example includes the two glutathione *S*-transferases (PGA1\_c13330 and \_c23760), which are involved in cellular detoxification by catalyzing the conjugation of glutathione with a wide range of endogenous and xenobiotic alkylating agents or by binding antibiotics and reduce the antimicrobial activity of beta-lactam drugs (51, 52). A fifth example is the homoserine/homoserine lactone efflux protein (PGA1\_c28370), which is part of the quorum-sensing system in roseobacters (53) and is suggested to be involved in antibiotic/exoenzyme production and biofilm formation in response to high population density (54). Finally, a gene encoding antibiotic biosynthesis monooxygenases (PGA1\_c18850) was also in the outlier list. Because *Phaeobacter* spp. are probiotic bacteria showing strong antagonistic interaction with the common aquaculture pathogens in the *Vibrionaceae* lineage (55–57), our finding suggests that allele acquisition from distant relatives is potentially an important mechanism supporting this major phenotype.

In another *Roseobacter* lineage, *Sulfitobacter pontiacus*, we identified seven gene families showing evidence of potential recombination occurring at the outgroup branch and another seven families showing evidence of potential recombination at ingroup branches (Fig. 4B; Fig. S1F). These outlier loci encode various biological functions, such as DNA transposition, signal transduction, and capsular polysaccharide synthesis, but they are not related to antibiotic production and degradation (Table S5B).

**Recombination with distant relatives in a marine SAR11 lineage.** The *k*-means procedure identified 55 genes showing unusual pattern of the  $d_s$  values in the five closely related genomes of "*Candidatus Pelagibacter ubique*" (see Fig. S4C and Table S6 in the supplemental material). In 8 of the 10 pairwise genome combinations, the mean and median values of  $d_s$  in this cluster were one or two orders of magnitude greater than those in the other cluster (Table S4C). We reconstructed the most likely scenarios of homologous recombination events that gave rise to the anomalous  $d_s$  at each of these 55 loci along the phylogeny of these five strains. The most frequently observed pattern (18 of 55) is that all pairs involved in HTCC1040 are highly divergent from each other at synonymous sites, while all remaining pairs are identical or nearly so. This can be readily explained by an acquisition of a divergent allele through recombination on the leaf branch of HTCC1040 or on the ancestral branch, giving rise to the most recent common ancestor (MRCA) of the other four strains (Fig. 5A). Another commonly observed pattern (6 of 55) is that all pairs involved in HTCC1016 have unusually high  $d_s$  values while the remaining pairs do not, suggesting that recombination occurred on the leaf branch of HTCC1016 (Fig. 5B). Recombination on other leaf branches (HTCC1002, HTCC1062, and HTCC1013) was rare (two gene families for HTCC1002 and one each for HTCC1013 and HTCC1062). For some loci (4 of 55), HTCC1002 and HTCC1062 are very similar to each other at synonymous sites, while they are highly divergent from HTCC1013, HTCC1016, and HTCC1040, suggesting that recombination occurred prior to the diversification of HTCC1002 and HTCC1062 (Fig. 5C). For other loci (1 of 55), HTCC1002, HTCC1062, and HTCC1013 are identical or nearly so to each other at synonymous sites, while they are highly divergent from HTCC1016 and HTCC1040. This can be easily explained by a recombination event having occurred prior to the diversification of HTCC1002, HTCC1062, and HTCC1013 (Fig. 5D). In the above cases, recombination most likely occurred only once. Among the genes showing these patterns include those encoding a number of substrate transporters, such as the TRAP-type  $C_4$ -dicarboxylate transporter (Ga0076703\_1168), tripartite ATP-independent periplasmic transporter (DctQ component) (Ga0076703\_1167), tripartite tricarboxylate transporter TctA family (Ga0076703\_1244), ABC transporters (mostly for sugar) (Ga0076703\_11155, \_11156, \_11157, and \_11158), permease for cytosine/purines (Ga0076703\_1164), mannitol transporter (Ga0076703\_1166), and  $Na^+$ -dependent transporter (Ga0076703\_11892) (Table S6).

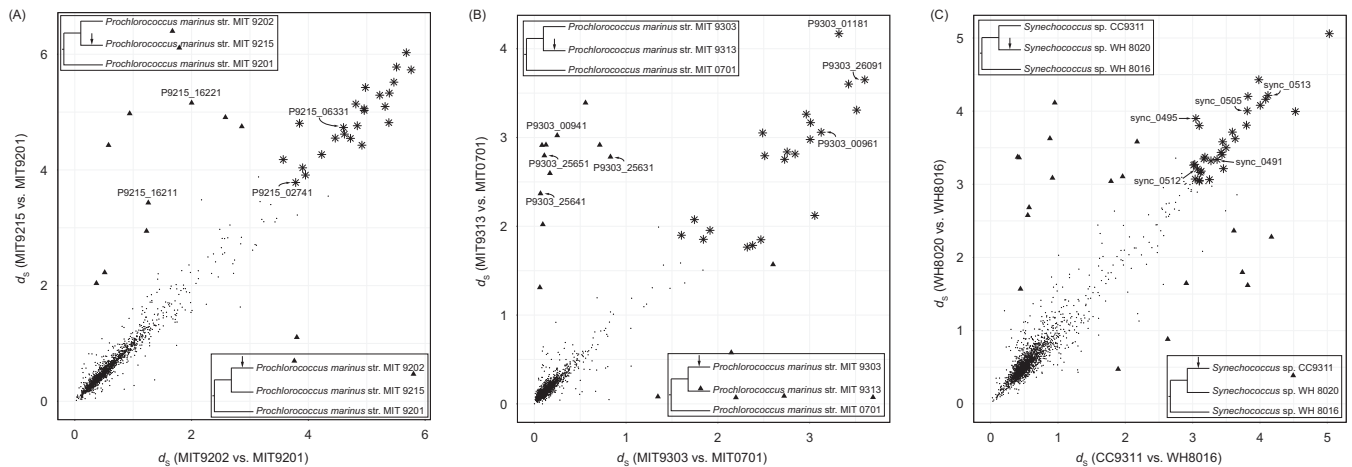
The remaining loci have more complex patterns of synonymous substitution, which must account for two or more recombination events. In two of these loci, for instance, HTCC1002 and HTCC1062 are identical at synonymous sites, while all other possible pairwise combinations are highly divergent. This requires at least three independent recombination events at three ancestral branches of the phylogeny (Fig. 5E). Interestingly, these loci encode an O-antigen ligase (Ga0076703\_11357) involved in the synthesis of O antigen, an LPS found in the outer membrane of Gram-negative bacteria. In another locus, encoding a putative lipoprotein (Ga0076703\_11990), all pairwise comparisons involving HTCC1062 or HTCC1013 are highly divergent at synonymous sites, but HTCC1062 and HTCC1013 are highly similar at synonymous sites, suggesting that either HTCC1062 or HTCC1013 acquired a divergent allele and this was followed by a recombination event between HTCC1062 and HTCC1013 (Fig. 5F). In a third locus, encoding a glycosyltransferase (Ga0076703\_12271) and an endonuclease (Ga0076703\_12272), all but the pair HTCC1016 and HTCC1062 are highly divergent at synonymous sites, suggesting that each of the three ancestral branches may have recombined with divergently related organisms and this was followed by a recombination event be-





**FIG 5** The phylogenetic relationships of five closely related *Candidatus Pelagibacter ubique* genomes, illustrating eight possible scenarios of recombination with distant relatives that gave rise to the anomalous  $d_5$  values at the 55 loci. This inference was based on the principle of parsimony. Cladograms, rather than phylograms, are used here in order to display the gene recombination among taxa appropriately.

tween HTCC1016 and HTCC1062 (Fig. 5G). In another six genes, the  $d_5$  values of all pairwise comparisons within the clade HTCC1002/HTCC1062/HTCC1013 are approximately one order of magnitude smaller than those of all the remaining pairwise comparisons, suggesting that recombination with a divergently related organism occurred on the ancestral branch leading to the MRCA of this clade and on the leaf branch of HTCC1040 or on the ancestral branch, giving rise to the clade HTCC1002/HTCC1062/HTCC1013/HTCC1016 (Fig. 5H). This pattern of recombination is evident in a gene cluster consisting of genes for CTP synthase (Ga0076703\_12347) and 2-dehydro-3-deoxy-phosphooctonate aldolase (Ga0076703\_12346), which is engaged in CMP-2-keto-3-deoxyoctulosonic acid (KDO) biosynthesis, a key step in the biosynthesis of LPS (58), as well as in another gene cluster involved in thiamine salvage comprising genes for a YggS-like protein with a TIM-barrel fold (Ga0076703\_11577), a thiamine monophosphate synthase (Ga0076703\_11578), and a putative porin (Ga0076703\_11579) (Table S6).



**FIG 6** Detection of  $d_s$  outliers of a high-light (HL)-adapted *Prochlorococcus* lineage (A), a low-light (LL)-adapted *Prochlorococcus* lineage (B), and *Synechococcus* 5.1B (C) through the linear regression procedure outlined in the Fig. 1 legend. Asterisks and solid triangles represent outlier genes in which the divergent alleles were acquired at the outgroup and ingroup branches, respectively. The predicted functions of outlier genes are given in Tables S7 and S8 in the supplemental material.

The recent isolation of phages (pelagiphage) from “*Candidatus Pelagibacter ubique*” and discovery of their high abundance in oceanic viromes suggest that SAR11 organisms interact with pelagiphage commonly in the ocean (59). In contrast to recombinational genes encoding various transporters affected by only one recombination event, multiple genes potentially involved in the interaction with phages were subject to repeated recombination with distant relatives, providing further evidence for the ecological significance of the SAR11-pelagiphage interaction. Among these, the gene encoding O-antigen ligase is located in the hypervariable region (HVR2), consistent with a previous speculation that HVR2 may be involved in phage resistance (59). Independent recombination events occurring at the O-antigen ligase gene strongly suggest that LPS plays an important but currently unrecognized role in SAR11 and perhaps many other marine bacteria.

**Recombination with distant relatives in marine *Prochlorococcus* lineages.** We also analyzed two high-light (HL)-adapted *Prochlorococcus* lineages (both in clade HLII) (Fig. 6A, S1G, S4D, and S5C; Table S4D) and two low-light (LL)-adapted *Prochlorococcus* lineages (clade LLIV and clade LLI) (Fig. 6B, Fig. S1H, S4E, and S5D; Table S4E) using the  $d_s$  outlier approaches. In the first HLII lineage, consisting of three genomes (MIT 9202, MIT 9201, and MIT 9215), 38 genes fall into the outlier group, where two genes coding for high-light-inducible proteins (HLIPs) (P9215\_16211 and \_16221) (see Table S7A in the supplemental material) show evidence for recombination with distant relatives. HLIPs play a critical role in protecting the cyanobacterial photosynthetic apparatus under stressful conditions, such as high-level light exposure, nitrogen/sulfur deprivation, and low temperature (60). These genes were found in *Prochlorococcus* phage genomes and are subject to repeated transfers among hosts through phage infection (61, 62). Phylogenetic analysis suggests that horizontal transfers for these two loci are potentially mediated by different phages (see Fig. S7 in the supplemental material). Another photosynthesis-related gene with evidence for recombination with distant relatives encodes the photosystem II (PSII) reaction center protein H (*psbH*; P9215\_02741) (Table S7A). This protein is potentially important for energy dissipation within PSII (63). We did not find *psbH* in the 78 available cyanophage genomes, and thus the mechanism underlying its transfer remains unknown. A further interesting finding was that a gene encoding carboxysome shell protein (P9215\_06331), involved in CO<sub>2</sub> fixation, is also part of the recombination gene pool. In contrast, in the second HLII lineage, consisting of four genomes (AS9601, MIT9301, MIT9314, and SB), only 20 genes show unusual patterns of  $d_s$  values. This list is enriched with genes involved in the

biosynthesis of the cell surface structure, including a UDP-glucose 6-dehydrogenase gene (A9601\_13901) and a cluster of four genes engaged in biosynthesis of GDP-mannose (A9601\_13971, \_13981, \_13991, and \_14001), a key substrate in glycoprotein formation (Table S7B).

In the case of the two LL lineages, genes involved in O-antigen biosynthesis are commonly found in the  $d_5$  outlier list in both lineages (Table S7C and D). Other outlier genes in the LLIV lineage encode dTDP-D-glucose 4,6-dehydratase (*rfbB*; P9303\_01181), NAD-dependent epimerase (P9303\_26091), dTDP-4-dehydrorhamnose 3,5-epimerase (*rfbC*; P9303\_00941), and dTDP-4-dehydrorhamnose reductase (*rfbD*; P9303\_00961). These together convert dTDP- $\alpha$ -D-glucose to  $\beta$ -L-rhamnopyranose (64–66), a deoxy-sugar that functions as a building block of the O antigens of many bacterial species (27). They also include several genes encoding group 1 glycosyltransferases (P9303\_25631, \_25641, and \_25651), which are engaged in cell envelope and outer membrane biogenesis. The outlier list in the LLI lineage includes genes for UDP-glucose 4-epimerase (NATL1\_08531) and a cluster catalyzing dTDP-L-rhamnose biosynthesis (NATL1\_08551, \_08561, and \_08571).

**Recombination with distant relatives in marine *Synechococcus* lineages.** We also applied the  $d_5$  outlier approaches to three lineages of *Synechococcus* affiliated with clade 5.1B (Fig. 6C), an unclassified clade (Fig. S1J), and clade 5.1A (Fig. S1K). For the 5.1B lineage, consisting of three genomes, the loci showing evidence for recombination at the outgroup branch feature a cluster of genes involved in the biosynthesis of phycobilisomes (PBS), which are accessory pigments in *Synechococcus* but are not visible in *Prochlorococcus* (67). The PBS component genes showing this unusual pattern of  $d_5$  include a gene encoding ferredoxin oxidoreductase (*sync\_0491*), which catalyzes the conversion of biliverdin-IX- $\alpha$  to phycoerythrobilin, *cpaB-1* and *cpaB-2* (*sync\_0495* and *\_0505*), which encode type 1 and 2 phycoerythrin pigments, and two PBS linker polypeptide genes (*sync\_0512* and *\_0513*), which ensure efficient energy transfer between the peripheral pigments and the photosynthetic reaction center (see Table S8A in the supplemental material). As the PBS pigment composition has been known as a niche-defining feature in *Synechococcus* (68, 69), the fine-tuning of the pigment-encoding sequences through homologous recombination with distant relatives is likely an important mechanism that *Synechococcus* uses to adapt to ocean niches where distinct wavelengths of light are available. In the other two *Synechococcus* lineages, photosynthesis-related genes are surprisingly depleted in the outlier list (Table S8B and C). In the unclassified lineage consisting of PCC7002 and NKBG042902, for instance, these outlier loci contain many genes potentially involved in interaction with phage, including genes for CRISPR-associated proteins Cas5 (SYNPCC7002\_D0021) and Cas2 (SYNPCC7002\_RS14945), phage integrase (SYNPCC7002\_A1888), glycosyltransferases (SYNPCC7002\_A1514 and \_A1517), and a polysaccharide biosynthesis export protein (SYNPCC7002\_A2451) involved in lipooligosaccharide biosynthesis (70). In the case of the 5.1A lineage, the outlier loci feature a number of genes for transporters, including a urea transporter (*Sync9605\_2617* and *\_2618*).

**Concluding remarks.** In sum, the 16 phylogenetically and ecologically divergent heterotrophic marine bacterial populations examined here all modified a subset of their gene repertoire through allele acquisition from evolutionarily distant relatives. This type of recombination may have distinctly facilitated substrate uptake, cell maintenance, phage resistance, antibiotic production, detoxification, adhesion, EPS production, hydrocarbon degradation, polysaccharide degradation, cold adaptation, photosynthesis, and CO<sub>2</sub> fixation, among other important cellular processes, and thus may represent an important way that bacteria rapidly adapt to a changing ocean. The  $d_5$  outlier approaches employed here allow for identification of important differences among closely related phylogenetic lineages (e.g., the two species in *Cellulophaga*, the high-light- and low-light-adapted ecotypes in *Prochlorococcus*, clades 5.1A and 5.1B in *Synechococcus*, and the two genera in the *Roseobacter* group) and among bacteria that are ecologically distinct but share nearly identical 16S rRNA genes (e.g., *Paraglaciecola chathamensis*

S18K6 from deep-sea sediment near hydrothermal vents compared to two strains from other sediments). As these bacteria are members of *Alphaproteobacteria*, *Gammaproteobacteria*, *Flavobacteria*, and *Cyanobacteria*, which dominate in global ocean bacterial communities (71–73), our finding has significant implications for microbial evolution and adaptation in the oligotrophic ocean.

## MATERIALS AND METHODS

**Genomic sequence download, ortholog identification, and phylogenetic analysis.** By the time this study was conducted, the genomic sequences varying at the strain level were available for multiple marine bacterial lineages (Table S1), and their predicted protein-coding sequences and 16S rRNA gene sequences were downloaded from NCBI RefSeq (74) and IMG/ER (75). Orthologous gene families shared by all members of a species and its outgroup were identified (Table S1) using the GET\_HOMOLOGUES package (76). Members in each gene family were aligned at the amino acid sequence level using the MAFFT software (77), and the columns with gaps were deleted. The trimmed alignments then were concatenated. Best-fitting partition schemes and substitution models were selected by PartitionFinder v2.1.1 (78). A phylogenomic tree for each species was constructed using the RAxML software (79) under the PROTGAMMA model with statistical support estimated from 100 bootstrapped replicates. Phylogenomic trees were also constructed using only the vertically derived genes by deleting the gene families whose members showed a significant signal of homologous recombination through the PhiTest (80) analysis. The resulting topologies are consistent across both methods. The branch lengths and support values obtained from the former method are used in the figures.

**Calculation of synonymous substitution rate.** All methods for the calculation of  $d_s$  assume a nucleotide evolution model, and these models fall in two categories depending on whether they take into account the base frequency. The Hasegawa-Kishino-Yano model, for instance, considers the base frequency, while the Kimura two-parameter model does not. Unrealistic results, in particular those for  $d_s$ , can be obtained if a model that does not control for the base frequency is applied to composition-biased DNA sequences (81). The KaKs\_Calculator software is designed to select the best-fit model that accounts for biased nucleotide content based on the corrected Akaike information criterion (82) and thus was employed in the present study, in which the genomic G+C content of the included taxa varies from below 30% to 60% (Table S1).

**Outlier identification.** We improved the approaches that were used to detect  $d_s$  outlier genes in a previous study (12). For a lineage with two closely related genomes (e.g., *Synechococcus* clade 5.1A; a complete list is in Table S1), we ranked the  $d_s$  values of each genome pair from least to greatest and plotted the  $d_s$  values against their ranks (Fig. S1J and K). A linear relationship between  $d_s$  and rank was tested with linear regression analysis (83). The Bonferroni correction was implemented to control for multiple comparisons using the `p.adjust` function of the “stats” base package of R (83). Loci showing statistically significant Studentized residuals after Bonferroni correction ( $P < 0.05$ ) were considered outliers.

For a lineage with three closely related genomes, one strain must be an outgroup of the other two strains. Recombination may have occurred in the outgroup strain (or the ancestral branch giving rise to the most recent common ancestor [MRCA] of the other two) and may have also occurred on the leaf branches of the other two. To differentiate these two types, different linear regression approaches were fitted to identify outliers. Taking *C. baltica* as an example, strain NN016038 is the outgroup of strains E6 and “18,” with the latter two considered “ingroup” strains (Fig. 1A; Fig. S1A). If a divergent allele was introduced into the outgroup (NN016038) or the ancestral branch giving rise to the MRCA of E6 and “18,” both E6 and “18” are expected to show unusually high  $d_s$  in comparison with NN016038 (Fig. S1A). If E6 acquired a divergent allele, the  $d_s$  of the E6/NN016038 pair at that locus is expected to be unusually high compared to the  $d_s$  of the “18”/NN016038 pair (Fig. 1A). Likewise, if strain “18” acquired a divergent allele, the  $d_s$  of the “18”/NN016038 pair at that locus should be unusually high in comparison to the  $d_s$  of the E6/NN016038 pair (Fig. 1A). To detect outliers of the first type, we ranked the  $d_s$  values of the single-copy orthologous genes shared by the E6/NN016038 pair from least to greatest (Fig. S1A) and performed a linear regression analysis of  $d_s$  against the rank. The same procedure was also applied to the “18”/NN016038 pair. Only those gene families identified as outliers in both genome pairs were considered loci subject to recombination occurring at outgroup (NN016038) or the ancestral branch giving rise to the MRCA of E6 and “18.” To detect outliers of the second type, we performed a linear regression analysis of the  $d_s$  calculated from the E6/NN016038 pair against the  $d_s$  calculated from the “18”/NN016038 pair after excluding loci identified as being of the first type. In both types, outliers were defined as gene families whose absolute Studentized residuals exceeded two times the standard deviation and remained significant after Bonferroni correction. The possible influence of outliers and leverage points upon the inference of linearity was minimized using a robust weighted-likelihood approach (`wle.lm` function in R) (83), which calibrates weights for observations based on their deviation from the smoothed kernel density (84).

For a lineage with four or more closely related genomes, we applied the *k*-means clustering procedure to identify genes showing unusual patterns of synonymous nucleotide substitution. Taking “*Candidatus* Pelagibacter ubique” as an example, this method used the all possible pairwise  $d_s$  values of “*Candidatus* Pelagibacter ubique” HTCC1002, HTCC1062, HTCC1013, HTCC1016, and HTCC1040 across 1,162 single-copy orthologous genes shared by these five genomes. The number (*k*) of clusters has to be provided before the *k*-means procedure can be started. The R package “NbClust” v2.0.1 (85) provides a wide variety of indices for cluster validity and thus was used to determine the optimal number of clusters.

For all the lineages under the  $k$ -means analysis,  $k$  was determined to be two as the optimal value (see Fig. S4A for *Marinobacter hydrocarbonoclasticus*, Fig. S4B for *Pseudoalteromonas haloplanktis*, Fig. S4C for “*Candidatus Pelagibacter ubique*,” Fig. S4D for *Prochlorococcus* HLII, and Fig. S4E for *Prochlorococcus* LLI).

**Syntenic block analysis.** Syntenic blocks of multiple genomes were identified using Sibelia (86), a sequence-referenced approach based on de Bruijn graphs. The cutoff for the minimum length of syntenic blocks was set as 5,000 bp, and 4 rounds of iteration were performed for fine-granularity scalable refinement. Genes in the orthologous families identified above were assigned to a syntenic block with regard to their loci on the chromosome/scaffold/contigs.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AEM.02545-17>.

**SUPPLEMENTAL FILE 1**, PDF file, 13.8 MB.

## ACKNOWLEDGMENTS

We thank Mary Ann Moran for providing valuable advice on an early version of the manuscript.

This research was funded by the Hong Kong Environment and Conservation Fund (15/2016), the Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16), and a Direct Grant (4930062) from the Chinese University of Hong Kong.

We declare no competing commercial interests in relation to this work.

## REFERENCES

- Dahlberg C, Bergström M, Hermansson M. 1998. In situ detection of high levels of horizontal plasmid transfer in marine bacterial communities. *Appl Environ Microbiol* 64:2670–2675.
- Jiang SC, Paul JH. 1998. Gene transfer by transduction in the marine environment. *Appl Environ Microbiol* 64:2780–2787.
- Rohwer F, Thurber RV. 2009. Viruses manipulate the marine environment. *Nature* 459:207–212. <https://doi.org/10.1038/nature08060>.
- Stewart GJ, Sinigalliano CD. 1990. Detection of horizontal gene transfer by natural transformation in native and introduced species of bacteria in marine and synthetic sediments. *Appl Environ Microbiol* 56:1818–1824.
- McDaniel LD, Young E, Delaney J, Ruhnan F, Ritchie KB, Paul JH. 2010. High frequency of horizontal gene transfer in the oceans. *Science* 330:50. <https://doi.org/10.1126/science.1192243>.
- Biller SJ, Schubotz F, Roggensack SE, Thompson AW, Summons RE, Chisholm SW. 2014. Bacterial vesicles in marine ecosystems. *Science* 343:183–186. <https://doi.org/10.1126/science.1243457>.
- Kolling GL, Matthews KR. 1999. Export of virulence genes and Shiga toxin by membrane vesicles of *Escherichia coli* O157:H7. *Appl Environ Microbiol* 65:3761–3761.
- Rumbo C, Fernandez-Moreira E, Merino M, Poza M, Mendez JA, Soares NC, Mosquera A, Chaves F, Bou G. 2011. Horizontal transfer of the OXA-24 carbapenemase gene via outer membrane vesicles: a new mechanism of dissemination of carbapenem resistance genes in *Acinetobacter baumannii*. *Antimicrob Agents Chemother* 55:3084–3090. <https://doi.org/10.1128/AAC.00929-10>.
- Luo H. 2015. The use of evolutionary approaches to understand single cell genomes. *Front Microbiol* 6:191. <https://doi.org/10.3389/fmicb.2015.00191>.
- Luo H, Swan BK, Stepanauskas R, Hughes AL, Moran MA. 2014. Comparing effective population sizes of dominant marine alphaproteobacteria lineages. *Environ Microbiol Rep* 6:167–172. <https://doi.org/10.1111/1758-2229.12129>.
- Hughes AL. 2008. Near neutrality: leading edge of the neutral theory of molecular evolution. *Ann N Y Acad Sci* 1133:162–179. <https://doi.org/10.1196/annals.1438.001>.
- Hughes AL, French JO. 2007. Homologous recombination and the pattern of nucleotide substitution in *Ehrlichia ruminantium*. *Gene* 387:31–37. <https://doi.org/10.1016/j.gene.2006.08.003>.
- Lin MZ, Kussell E. 2017. Correlated mutations and homologous recombination within bacterial populations. *Genetics* 205:891–917. <https://doi.org/10.1534/genetics.116.189621>.
- Vos M. 2009. Why do bacteria engage in homologous recombination? *Trends Microbiol* 17:226–232. <https://doi.org/10.1016/j.tim.2009.03.001>.
- Engel P, Stepanauskas R, Moran NA. 2014. Hidden diversity in honey bee gut symbionts detected by single-cell genomics. *PLoS Genet* 10:e1004596. <https://doi.org/10.1371/journal.pgen.1004596>.
- Hellweger FL, Huang Y, Luo H. 2018. Carbon limitation drives GC content evolution of a marine bacterium in an individual-based genome-scale model. *ISME J* <https://doi.org/10.1038/s41396-017-0023-7>.
- Luo H, Thompson LR, Stingl U, Hughes AL. 2015. Selection maintains low genomic GC content in marine SAR11 lineages. *Mol Biol Evol* 32:2738–2748. <https://doi.org/10.1093/molbev/msv149>.
- Hughes AL, Friedman R. 2005. Nucleotide substitution and recombination at orthologous loci in *Staphylococcus aureus*. *J Bacteriol* 187:2698–2704. <https://doi.org/10.1128/JB.187.8.2698-2704.2005>.
- Hehemann JH, Boraston AB, Czjzek M. 2014. A sweet new wave: structures and mechanisms of enzymes that digest polysaccharides from marine algae. *Curr Opin Struct Biol* 28:77–86. <https://doi.org/10.1016/j.sbi.2014.07.009>.
- Martens EC, Lowe EC, Chiang H, Pudlo NA, Wu M, McNulty NP, Abbott DW, Henrissat B, Gilbert HJ, Bolam DN, Gordon JL. 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* 9:e1001221. <https://doi.org/10.1371/journal.pbio.1001221>.
- Martens EC, Koropatkin NM, Smith TJ, Gordon JL. 2009. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* 284:24673–24677. <https://doi.org/10.1074/jbc.R109.022848>.
- López-Pérez M, Rodríguez-Valera F. 2016. Pangenome evolution in the marine bacterium *Alteromonas*. *Genome Biol Evol* 8:1556–1570. <https://doi.org/10.1093/gbe/evw098>.
- López-Pérez M, Gonzaga A, Martín-Cuadrado A-B, Onyshchenko O, Ghavidel A, Ghai R, Rodríguez-Valera F. 2012. Genomes of surface isolates of *Alteromonas macleodii*: the life of a widespread marine opportunistic copiotroph. *Sci Rep* 2:696. <https://doi.org/10.1038/srep00696>.
- Nwodo UU, Green E, Okoh AI. 2012. Bacterial exopolysaccharides: functionality and prospects. *Int J Mol Sci* 13:14002–14015. <https://doi.org/10.3390/ijms131114002>.
- Mehta A, Sidhu C, Pinnaka AK, Roy Choudhury A. 2014. Extracellular polysaccharide production by a novel osmotolerant marine strain of *Alteromonas macleodii* and its application towards biomineralization of silver. *PLoS One* 9:e98798. <https://doi.org/10.1371/journal.pone.0098798>.
- Zhang Z, Li Z, Jiao N. 2014. Effects of D-amino acids on the EPS production and cell aggregation of *Alteromonas macleodii* strain JL2069. *Curr Microbiol* 68:751–755. <https://doi.org/10.1007/s00284-014-0520-0>.
- Samuel G, Reeves P. 2003. Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* 338:2503–2519. <https://doi.org/10.1016/j.carres.2003.07.009>.

28. Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D. 2011. Genomic island variability facilitates *Prochlorococcus*-virus coexistence. *Nature* 474: 604–608. <https://doi.org/10.1038/nature10172>.
29. Goon S, Kelly JF, Logan SM, Ewing CP, Guerry P. 2003. Pseudaminic acid, the major modification on *Campylobacter* flagellin, is synthesized via the Cj1293 gene. *Mol Microbiol* 50:659–671. <https://doi.org/10.1046/j.1365-2958.2003.03725.x>.
30. Zhilenkov EL, Popova VM, Popov DV, Zavalsky LY, Svetoch EA, Stern NJ, Seal BS. 2006. The ability of flagellum-specific *Proteus vulgaris* bacteriophage PV22 to interact with *Campylobacter jejuni* flagella in culture. *Virology* 3:50. <https://doi.org/10.1186/1743-422X-3-50>.
31. Matsuyama H, Hirabayashi T, Kasahara H, Minami H, Hoshino T, Yumoto I. 2006. *Glaciecola chathamensis* sp. nov., a novel marine polysaccharide-producing bacterium. *Int J Syst Evol Microbiol* 56:2883–2886. <https://doi.org/10.1099/ijs.0.64413-0>.
32. Thompson R. 1991. Gazetteer of seafloor features in the New Zealand region. New Zealand Oceanographic Institute, Wellington, New Zealand.
33. Vetrani C, Chew YS, Miller SM, Yagi J, Coombs J, Lutz RA, Barkay T. 2005. Mercury adaptation among bacteria from a deep-sea hydrothermal vent. *Appl Environ Microbiol* 71:220–226. <https://doi.org/10.1128/AEM.71.1.220-226.2005>.
34. Sinaei M, Rahmanpour S. 2013. Evaluation of glutathione S-transferase activity as a biomarker of PAH pollution in mudskipper, *Boleophthalmus dussumieri*, Persian Gulf. *Bull Environ Contam Toxicol* 90:369–374. <https://doi.org/10.1007/s00128-012-0917-z>.
35. Overholt WA, Green SJ, Marks KP, Venkatraman R, Prakash O, Kostka JE. 2013. Draft genome sequences for oil-degrading bacterial strains from beach sands impacted by the Deepwater Horizon oil spill. *Genome Announc* 1:e01015-13. <https://doi.org/10.1128/genomeA.01015-13>.
36. Grimaud R, Ghiglione JF, Cagnon C, Lauga B, Vaysse PJ, Rodriguez-Blanco A, Mangenot S, Cruveiller S, Barbe V, Duran R, Wu LF, Talla E, Bonin P, Michotey V. 2012. Genome sequence of the marine bacterium *Marinobacter hydrocarbonoclasticus* SP17, which forms biofilms on hydrophobic organic compounds. *J Bacteriol* 194:3539–3540. <https://doi.org/10.1128/JB.00500-12>.
37. Mounier J, Camus A, Mitteau I, Vaysse PJ, Goulas P, Grimaud R, Sivadon P. 2014. The marine bacterium *Marinobacter hydrocarbonoclasticus* SP17 degrades a wide range of lipids and hydrocarbons through the formation of oleolytic biofilms with distinct gene expression profiles. *FEMS Microbiol Ecol* 90:816–831. <https://doi.org/10.1111/1574-6941.12439>.
38. Lamendella R, Strutt S, Borglin S, Chakraborty R, Tas N, Mason OU, Hultman J, Prestat E, Hazen TC, Jansson JK. 2014. Assessment of the Deepwater Horizon oil spill impact on Gulf coast microbial communities. *Front Microbiol* 5:130. <https://doi.org/10.3389/fmicb.2014.00130>.
39. Alonso-Gutierrez J, Figueras A, Albaiges J, Jimenez N, Vinas M, Solanas AM, Novoa B. 2009. Bacterial communities from shoreline environments (Costa da Morte, northwestern Spain) affected by the *Prestige* oil spill. *Appl Environ Microbiol* 75:3407–3418. <https://doi.org/10.1128/AEM.01776-08>.
40. Alterio V, Aurilia V, Romanelli A, Parracino A, Saviano M, D'Auria S, De Simone G. 2010. Crystal structure of an S-formylglutathione hydrolase from *Pseudoalteromonas haloplanktis* TAC125. *Biopolymers* 93:669–677. <https://doi.org/10.1002/bip.21420>.
41. Medigue C, Krin E, Pascal G, Barbe V, Bernsel A, Bertin PN, Cheung F, Cruveiller S, D'Amico S, Duilio A, Fang G, Feller G, Ho C, Mangenot S, Marino G, Nilsson J, Parrilli E, Rocha EP, Rouy Z, Sekowska A, Tutino ML, Vallenet D, von Heijne G, Danchin A. 2005. Coping with cold: the genome of the versatile marine Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Res* 15:1325–1335. <https://doi.org/10.1101/gr.4126905>.
42. Aliyu H, De Maayer P, Cowan D. 2016. The genome of the Antarctic polyextremophile *Nesterenkonia* sp. AN1 reveals adaptive strategies for survival under multiple stress conditions. *FEMS Microbiol Ecol* 92:fiw032. <https://doi.org/10.1093/femsec/fiw032>.
43. Aussel L, Pierrrel F, Loiseau L, Lombard M, Fontecave M, Barras F. 2014. Biosynthesis and physiology of coenzyme Q in bacteria. *Biochim Biophys Acta* 1837:1004–1011. <https://doi.org/10.1016/j.bbabi.2014.01.015>.
44. Hoque MA, Uraji M, Banu MN, Mori IC, Nakamura Y, Murata Y. 2010. The effects of methylglyoxal on glutathione S-transferase from *Nicotiana tabacum*. *Biosci Biotechnol Biochem* 74:2124–2126. <https://doi.org/10.1271/bbb.100393>.
45. Linhartova I, Bumba L, Masin J, Basler M, Osicka R, Kamanova J, Prochazkova K, Adkins I, Hejnova-Holubova J, Sadilkova L, Morova J, Sebo P. 2010. RTX proteins: a highly diverse family secreted by a common mechanism. *FEMS Microbiol Rev* 34:1076–1112. <https://doi.org/10.1111/j.1574-6976.2010.00231.x>.
46. Inoue Y, Kimura A. 1995. Methylglyoxal and regulation of its metabolism in microorganisms. *Adv Microb Physiol* 37:177–227. [https://doi.org/10.1016/S0065-2911\(08\)60146-0](https://doi.org/10.1016/S0065-2911(08)60146-0).
47. Korithoski B, Levesque CM, Cvitkovitch DG. 2007. Involvement of the detoxifying enzyme lactoylglutathione lyase in *Streptococcus mutans* aciduricity. *J Bacteriol* 189:7586–7592. <https://doi.org/10.1128/JB.00754-07>.
48. Daigle DM, McKay GA, Wright GD. 1997. Inhibition of aminoglycoside antibiotic resistance enzymes by protein kinase inhibitors. *J Biol Chem* 272:24755–24758. <https://doi.org/10.1074/jbc.272.40.24755>.
49. Cai Y, Trodler P, Jiang S, Zhang W, Wu Y, Lu Y, Yang S, Jiang W. 2009. Isolation and molecular characterization of a novel D-hydantoinase from *Jannaschia* sp. CCS1. *FEBS J* 276:3575–3588. <https://doi.org/10.1111/j.1742-4658.2009.07077.x>.
50. Trent MS, Pabich W, Raetz CR, Miller SI. 2001. A PhoP/PhoQ-induced lipase (PagL) that catalyzes 3-O-deacylation of lipid A precursors in membranes of *Salmonella typhimurium*. *J Biol Chem* 276:9083–9092. <https://doi.org/10.1074/jbc.M010730200>.
51. Jakoby WB. 1978. The glutathione S-transferases: a group of multifunctional detoxification proteins. *Adv Enzymol Relat Areas Mol Biol* 46: 383–414.
52. Sheehan D, Meade G, Foley VM, Dowd CA. 2001. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem J* 360:1–16.
53. Cude WN, Buchan A. 2013. Acyl-homoserine lactone-based quorum sensing in the Roseobacter clade: complex cell-to-cell communication controls multiple physiologies. *Front Microbiol* 4:336. <https://doi.org/10.3389/fmicb.2013.00336>.
54. Gram L, Grossart HP, Schlingloff A, Kiorboe T. 2002. Possible quorum sensing in marine snow bacteria: production of acylated homoserine lactones by Roseobacter strains isolated from marine snow. *Appl Environ Microbiol* 68:4111–4116. <https://doi.org/10.1128/AEM.68.8.4111-4116.2002>.
55. D'Alvise PW, Lillebø S, Prol-Garcia MJ, Wergeland HI, Nielsen KF, Bergh Ø, Gram L. 2012. *Phaeobacter gallaeciensis* reduces *Vibrio anguillarum* in cultures of microalgae and rotifers, and prevents vibriosis in cod larvae. *PLoS One* 7:e43996. <https://doi.org/10.1371/journal.pone.0043996>.
56. Prol-García M, Pintado J. 2013. Effectiveness of probiotic *Phaeobacter* bacteria grown in biofilters against *Vibrio anguillarum* infections in the rearing of turbot (*Psetta maxima*) larvae. *Mar Biotechnol* (NY) 15: 726–738. <https://doi.org/10.1007/s10126-013-9521-4>.
57. Prol García MJ, D'Alvise PW, Gram L. 2013. Disruption of cell-to-cell signaling does not abolish the antagonism of *Phaeobacter gallaeciensis* toward the fish pathogen *Vibrio anguillarum* in algal systems. *Appl Environ Microbiol* 79:5414–5417. <https://doi.org/10.1128/AEM.01436-13>.
58. Gattis SG, Chung HS, Trent MS, Raetz CR. 2013. The origin of 8-amino-3,8-dideoxy-D-manno-octulosonic acid (Kdo8N) in the lipopolysaccharide of *Shewanella oneidensis*. *J Biol Chem* 288:9216–9225. <https://doi.org/10.1074/jbc.M113.453324>.
59. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, Ellisman M, Deerinck T, Sullivan MB, Giovannoni SJ. 2013. Abundant SAR11 viruses in the ocean. *Nature* 494:357–360. <https://doi.org/10.1038/nature11921>.
60. He Q, Dolganov N, Bjorkman O, Grossman AR. 2001. The high light-inducible polypeptides in *Synechocystis* PCC6803. Expression and function in high light. *J Biol Chem* 276:306–314.
61. Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. 2004. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci U S A* 101:11013–11018. <https://doi.org/10.1073/pnas.0401526101>.
62. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. 2005. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 438:86–89. <https://doi.org/10.1038/nature04111>.
63. D'Haese SE, Sobotka R, Bucinska L, Dekker JP, Komenda J. 2015. Interaction of the PsbH subunit with a chlorophyll bound to histidine 114 of CP47 is responsible for the red 77K fluorescence of photosystem II. *Biochim Biophys Acta* 1847:1327–1334. <https://doi.org/10.1016/j.bbabi.2015.07.003>.
64. Giraud MF, McMiken HJ, Leonard GA, Messner P, Whitfield C, Naismith JH. 1999. Overexpression, purification, crystallization and preliminary structural study of dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD), the fourth enzyme of the dTDP-L-rhamnose synthesis pathway, from

- Salmonella enterica* serovar Typhimurium. Acta Crystallogr D Biol Crystallogr 55:2043–2046. <https://doi.org/10.1107/S0907444999012251>.
65. Allard ST, Giraud MF, Whitfield C, Graninger M, Messner P, Naismith JH. 2001. The crystal structure of dTDP-D-glucose 4,6-dehydratase (RmlB) from *Salmonella enterica* serovar Typhimurium, the second enzyme in the dTDP-L-rhamnose pathway. J Mol Biol 307:283–295. <https://doi.org/10.1006/jmbi.2000.4470>.
  66. Giraud MF, Leonard GA, Field RA, Berling C, Naismith JH. 2000. RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. Nat Struct Biol 7:398–402. <https://doi.org/10.1038/75178>.
  67. Partensky F, Hess WR, Vaulot D. 1999. *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. Microbiol Mol Biol Rev 63:106–127.
  68. Six C, Thomas JC, Garczarek L, Ostrowski M, Dufresne A, Blot N, Scanlan DJ, Partensky F. 2007. Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. Genome Biol 8:R259. <https://doi.org/10.1186/gb-2007-8-12-r259>.
  69. Haverkamp TH, Schouten D, Doeleman M, Wollenzien U, Huisman J, Stal LJ. 2009. Colorful microdiversity of *Synechococcus* strains (picocyanobacteria) isolated from the Baltic Sea. ISME J 3:397–408. <https://doi.org/10.1038/ismej.2008.118>.
  70. Stoddard LI, Martiny JB, Marston MF. 2007. Selection and characterization of cyanophage resistance in marine *Synechococcus* strains. Appl Environ Microbiol 73:5516–5522. <https://doi.org/10.1128/AEM.00356-07>.
  71. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooshep S, Wu D, Eisen JA, Hoffman JM, Remington K, Beeson K, Tran B, Smith H, Baden-Tillson H, Stewart C, Thorpe J, Freeman J, Andrews-Pfannkoch C, Venter JE, Li K, Kravitz S, Heidelberg JF, Utterback T, Rogers Y-H, Falcn LI, Souza V, Bonilla-Rosso G, Eguiarte LE, Karl DM, Sathyendranath S, Platt T, Bermingham E, Gallardo V, Tamayo-Castillo G, Ferrari MR, Strausberg RL, Nealson K, Friedman R, Frazier M, Venter JC. 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol 5:e77. <https://doi.org/10.1371/journal.pbio.0050077>.
  72. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature 345:60–63. <https://doi.org/10.1038/345060a0>.
  73. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Comejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmiento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Tara Oceans coordinators, Bowler TOC, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, et al. 2015. Structure and function of the global ocean microbiome. Science <https://doi.org/10.1126/science.1261359>.
  74. Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:D61–D65. <https://doi.org/10.1093/nar/gkl842>.
  75. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res 40:D115–D122. <https://doi.org/10.1093/nar/gkr1044>.
  76. Contreras-Moreira B, Vinuesa P. 2013. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol 79:7696–7701. <https://doi.org/10.1128/AEM.02411-13>.
  77. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
  78. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. Partition-Finder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol 34:772–773. <https://doi.org/10.1093/molbev/msw260>.
  79. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
  80. Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172:2665–2681. <https://doi.org/10.1534/genetics.105.048975>.
  81. Luo H, Hughes AL. 2012. dN/dS does not show positive selection drives separation of polar-tropical SAR11 populations. Mol Syst Biol 8:625. <https://doi.org/10.1038/msb.2012.58>.
  82. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. 2006. KaKs\_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinformatics 4:259–263. [https://doi.org/10.1016/S1672-0229\(07\)60007-2](https://doi.org/10.1016/S1672-0229(07)60007-2).
  83. R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
  84. Agostinelli C, Markatou M. 1998. A one-step robust estimator for regression based on the weighted likelihood reweighting scheme. Stat Prob Lett 37:341–350. [https://doi.org/10.1016/S0167-7152\(97\)00136-3](https://doi.org/10.1016/S0167-7152(97)00136-3).
  85. Charrad M, Ghazzali N, Boiteau V, Niknafs A. 2014. Nbclust: an R package for determining the relevant number of clusters in a data set. J Stat Softw 61:1–36. <https://doi.org/10.18637/jss.v061.i06>.
  86. Minkin I, Patel A, Kolmogorov M, Vyahhi N, Pham S. 2013. Sibelia: a scalable and comprehensive synteny block generation tool for closely related microbial genomes, p 215–229. In Darling A, Stoye J (ed), Algorithms in bioinformatics. WABI 2013. Lecture notes in computer science, vol 8126. Springer, Berlin, Germany.