AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE

Comparative Genomics of Cyclic di-GMP Metabolism and Chemosensory Pathways in *Shewanella algae* Strains: Novel Bacterial Sensory Domains and Functional Insights into Lifestyle Regulation

Alberto J. Martín-Rodríguez,[a] Shawn M. Higdon,[a] Kaisa Thorell,[b,c] Christian Tellgren-Roth,[d] Åsa Sjöling,[a] Michael Y. Galperin,[e] Tino Krell,[f] Ute Römling[a]

[a]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden
[b]Institute of Biomedicine, Department of Infectious Diseases, University of Gothenburg, Gothenburg, Sweden
[c]Västra Götaland Region, Department of Clinical Microbiology, Sahlgrenska University Hospital, Gothenburg, Sweden
[d]Department of Immunology, Genetics and Pathology, Uppsala Genome Center, Uppsala University, Uppsala, Sweden
[e]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA
[f]Department of Environmental Protection, Estación Experimental del Zaidín, Spanish National Research Council (CSIC), Granada, Spain

ABSTRACT *Shewanella* spp. play important ecological and biogeochemical roles, due in part to their versatile metabolism and swift integration of stimuli. While *Shewanella* spp. are primarily considered environmental microbes, *Shewanella algae* is increasingly recognized as an occasional human pathogen. *S. algae* shares the broad metabolic and respiratory repertoire of *Shewanella* spp. and thrives in similar ecological niches. In *S. algae*, nitrate and dimethyl sulfoxide (DMSO) respiration promote biofilm formation strain specifically, with potential implication of taxis and cyclic diguanosine monophosphate (c-di-GMP) signaling. Signal transduction systems in *S. algae* have not been investigated. To fill these knowledge gaps, we provide here an inventory of the c-di-GMP turnover proteome and chemosensory networks of the type strain *S. algae* CECT 5071 and compare them with those of 41 whole-genome-sequenced clinical and environmental *S. algae* isolates. Besides comparative analysis of genetic content and identification of laterally transferred genes, the occurrence and topology of c-di-GMP turnover proteins and chemoreceptors were analyzed. We found *S. algae* strains to encode 61 to 67 c-di-GMP turnover proteins and 28 to 31 chemoreceptors, placing *S. algae* near the top in terms of these signaling capacities per Mbp of genome. Most c-di-GMP turnover proteins were predicted to be catalytically active; we describe in them six novel N-terminal sensory domains that appear to control their catalytic activity. Overall, our work defines the c-di-GMP and chemosensory signal transduction pathways in *S. algae*, contributing to a better understanding of its ecophysiology and establishing *S. algae* as an auspicious model for the analysis of metabolic and signaling pathways within the genus *Shewanella*.

IMPORTANCE *Shewanella* spp. are widespread aquatic bacteria that include the well-studied freshwater model strain *Shewanella oneidensis* MR-1. In contrast, the physiology of the marine and occasionally pathogenic species *Shewanella algae* is poorly understood. Chemosensory and c-di-GMP signal transduction systems integrate environmental stimuli to modulate gene expression, including the switch from a planktonic to sessile lifestyle and pathogenicity. Here, we systematically dissect the c-di-GMP proteome and chemosensory pathways of the type strain *S. algae* CECT 5071 and 41 additional *S. algae* isolates. We provide insights into the activity and function of these proteins, including a description of six novel sensory domains. Our work will enable future analyses of the complex, intertwined c-di-GMP metabolism and chemotaxis networks of *S. algae* and their ecophysiological role.

Editor Ákos T. Kovács, Technical University of Denmark

Address correspondence to Alberto J. Martín-Rodríguez, jonatan.martin.rodriguez@ki.se.

The authors declare no conflict of interest.

Received 5 January 2022
Accepted 26 February 2022
Published 21 March 2022

KEYWORDS c-di-GMP, chemotaxis, sensing, signal transduction, whole-genome sequencing, *Shewanella*

The gammaproteobacterial genus *Shewanella* comprises more than 70 species of facultative anaerobes that thrive in aquatic ecosystems, such as the water column, sedimental microbial communities, the microbiota of animals, and biofilms (1, 2). The renowned physiological versatility of *Shewanella* spp. is reflected by their isolation from habitats with remarkably diverse environmental conditions with respect to temperature, from polar to tropical (3, 4); salinity, including hypersaline environments (5); pressure, including deep-sea environments (6); and oxygen concentration, including hypoxic or anoxic waters (7). Their widespread distribution and outstanding metabolic toolset make *Shewanella* spp. important players in global biogeochemical cycles, and significant research and technical efforts are devoted to exploiting their biotechnological potential (1). While most *Shewanella* spp. are regarded as environmental bacteria, there is increasing evidence of the pathogenic potential in some species. The species *Shewanella algae* is responsible for approximately 80% of all *Shewanella* infections in humans (8). In the majority of cases, patients have underlying conditions, but occasionally, healthy individuals are affected (8, 9). Besides, *S. algae* has been reported to be the causative agent of animal infections, including mass mortality events of reared abalone (10) and ulcer disease in reared fish (11, 12), which underscores its pathogenic potential to higher eukaryotic organisms.

Prokaryotic genome plasticity contributes to niche adaptation, which can be reflected at the genomic level by a variable accessory genome content, single nucleotide polymorphisms, and genome rearrangements. Strain-specific adaptation involving gene acquisition or loss has been described for a representative set of *Shewanella* spp. (13). Horizontal gene transfer has been recognized as an important driving force shaping *Shewanella* genomes, including genes involved in metal reduction (7, 14). We have previously reported strain-specific, respiration-driven biofilm formation in *S. algae* (15), a phenotype and potential niche adaptation mechanism in which sensing and signal transduction systems may play a role.

The switch from a free-living, planktonic lifestyle to a sessile, biofilm mode of life constitutes a major physiological challenge for bacteria. In most Gram-negatives, the second messenger cyclic diguanosine monophosphate (c-di-GMP) regulates this lifestyle transition. Thus, high (local) c-di-GMP levels can correlate with a biofilm phenotype, and low (local) c-di-GMP pools are frequently associated with motile behavior (16, 17). Biosynthesis of c-di-GMP from GTP is performed by GGDEF domain-containing proteins termed diguanylate cyclases (DGCs), whereas its breakdown into GMP and pGpG is catalyzed by c-di-GMP-specific phosphodiesterases (PDEs) through either EAL or HD-GYP domains (16, 17). Dual GGDEF-EAL domain-containing proteins can possess either DGC or PDE activity, or both, or be catalytically inactive (16, 17). Most c-di-GMP turnover proteins harbor N-terminal sensory input domains that respond to environmental stimuli to control catalytic activities of the output domains (18). Local, target-specific cytosolic pools of c-di-GMP are increasingly recognized to play important regulatory roles (19). An example is c-di-GMP production by AdrA (DgcC) in Salmonella enterica serovar Typhimurium and Escherichia coli, which largely acts in situ to activate the cellulose synthase BcsA (20, 21). Shewanella spp. are among the bacteria with the highest number and density of GGDEF, EAL, and HD-GYP domain-encoding genes in their genomes (https://www.ncbi.nlm.nih.gov/Complete_Genomes/c-di-GMP.html), which represents the c-di-GMP proportion of the "bacterial IQ" (22, 23). This suggests a significant impact of c-di-GMP signaling on the physiology of *Shewanella* spp. (24–26), including *S. algae* (15), which is also observed in other aquatic gammaproteobacteria such as *Vibrio* spp. (27, 28).

The complex physiology of *Shewanella* spp. requires fine-tuned integration of environmental stimuli with regulation at all levels from gene expression to posttranslational modifications. Chemosensory pathways represent major bacterial signal transduction systems that can be classified into three main families. Members of the Fla family

mediate flagellum-based chemotaxis, those of the TFP family are associated with type IV pilus-based motility, whereas the ACF family systems carry out alternative cellular functions like the control of c-di-GMP levels (29). Chemotactic movements are most frequently chemoattraction (i.e., movements to more optimal conditions for growth or survival) whereas there are relatively few cases of chemorepulsion (30). Energy taxis is an alternative mode of flagellum-based swimming motility that results in migration toward environments with more optimal levels of metabolic resources (30).

In a canonical chemotaxis chemosensory pathway, chemoeffectors bind to the extracytosolic ligand-binding domains of chemoreceptors, an event that triggers a molecular stimulus that is transmitted across the membrane, where it modulates the activity of the CheA autokinase, which in turn modulates the phosphorylation of the response regulator CheY. Only the phosphorylated form of CheY binds to the flagellar motor, allowing chemotaxis to occur. In addition, the sensitivity of chemosensory pathways is adjusted to the chemoeffector concentration by the concerted action of the methyltransferase CheR and methylesterase CheB, which control the methylation state of several glutamate residues in the chemoreceptor's signaling domain (31, 32). In addition, chemosensory pathways employ a number of auxiliary proteins that are present in some, but not all, pathways, such as the CheY phosphatases CheC, CheX, and CheZ (29).

Chemotaxis has been extensively studied in the model species *E. coli*. However, over the last decades chemotaxis has also been studied in other species with a different lifestyle, revealing a wide diversity in the number and type of chemoreceptors as well as in the signaling mechanisms (33, 34). *S. algae*, inhabiting environmental and host-associated niches with different levels of trophism, constitutes an attractive archetypal model for the understanding of *Shewanella* species chemotactic systems and their involvement in niche colonization, evasion, and pathogenesis. Chemotaxis of marine bacteria is currently poorly investigated, revealing a need for further research in this area. There is evidence that *Shewanella* spp. perform chemotaxis (35, 36) and energy taxis (37). In addition, there is initial evidence for an additional chemosensory pathway in *Shewanella* (38) that is currently of unknown function.

In this work, we define the repertoire of c-di-GMP turnover proteins and chemosensory systems in a collection of 42 *S. algae* strains, including the type strain, CECT 5071^T (=DSM 9167^T), whose complete genome sequence has been recently obtained (39). Through whole-genome sequencing (WGS), that included closing of 4 additional genomes, and subsequent bioinformatic analysis, we analyzed the pangenome of this strain collection and identified a diverse set of accessory genes that suggest a significant extent of physiological heterogeneity. We describe here the reference c-di-GMP proteome of the type strain *S. algae* CECT 5071, which, according to the presence and distribution of key signature motifs, consists of predicted catalytically active and inactive members. Reconstruction of the GGDEF, EAL, and HD-GYP domain phylogenies revealed a diverse origin of the members of this proteome, and intraspecific comparisons showcased substantial variability among isolates. Analysis of the N-terminal regions of c-di-GMP turnover proteins unraveled six novel bacterial sensory domains, including a previously unrecognized variant of the CSS domain. Furthermore, we dissect the chemosensory pathways of these 42 strains, including chemoreceptor topology, diversity, and gene synteny. Altogether, our work provides a comprehensive analysis aimed toward disentangling the complex and intertwined networks regulating motile and biofilm behaviors in this bacterial species and their ecophysiological implications.

RESULTS AND DISCUSSION

Pangenome analysis reveals distinct accessory genes in genetically independent strain backgrounds. We initiated our study by analyzing genome sequence characteristics and strain assignment to the species *S. algae*, since misidentification of *Shewanella* sp. isolates has been documented, including the reclassification of *Shewanella upenei* (strain CCUG 58400 in our set) and *Shewanella haliotis* as later heterotypic synonyms of *S. algae* (40–42). The identity of the isolates *as S. algae* was confirmed by multilocus

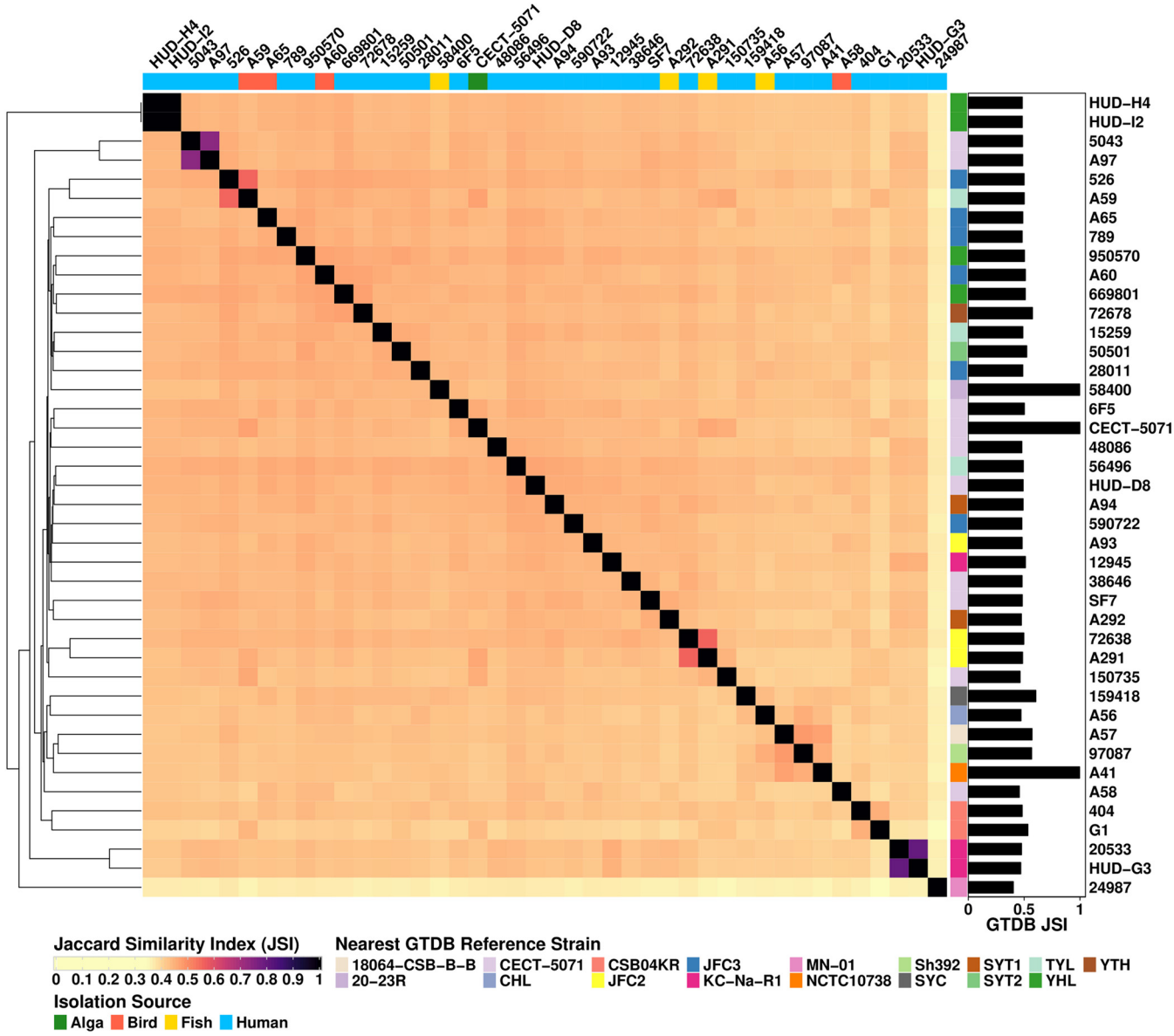

FIG 1 WGS-based distance and taxonomic comparisons using MinHash. An all-by-all pairwise comparison of the MinHash genome sketch for each *S. algae* strain. The clustered distance matrix depicts the JSI, also referred to as Jaccard distance, for each unique strain pairing based on comparisons of WGS-assembly nucleotide composition. JSI values closer to 1 (darker) showcase higher WGS similarity, while values approaching 0 (lighter) reflect greater dissimilarity. The column annotation bar indicates each strain's isolation source. Row annotation box colors reflect the *S. algae* reference genome from the GTDB with the highest JSI value to each of the 42 strains included in the study. Row annotation bar plot values represent the JSI between the query and GTDB subject strains.

sequence typing (MLST) (41) and pairwise digital DNA-DNA hybridization (dDDH) (43; data not shown). All strains belong to the same subspecies, as determined by a dDDH >79% cutoff (43; data not shown). *S. algae* genomes ranged from 4.66 to 5.07 Mbp, with a GC content in the range from 52.8% to 53.2%, which agreed with reported data for this species (40).

To evaluate the WGS diversity of the 42 *S. algae* isolates at the nucleotide composition level, we calculated the genomic distance, measured as the Jaccard Similarity Index (JSI), between each unique isolate pairing with MinHash sketches of their respective WGS assemblies (Fig. 1). The JSI for each of the all-by-all comparisons conducted spanned a range from 0.34 to 1, where lower values confirmed greater genomic distance and a value of 1 indicated absolute genomic similarity. Queries of all 42 MinHash genome sketches against the Genome Taxonomy Database (GTDB) corroborated the

dDDH and MLST results by producing an array of 17 best-matching reference genomes with an average query to subject JSI of 0.53. These queries included three sequenced isolates that had a 100% match to *S. algae* GTDB references, namely, the type strain CECT 5071, CCUG 58400 (originally reported as *S. upenei* 20-23R^T and later as a heterotypic synonym of *S. algae* as indicated above), and A41 (strain designation equivalent to NCTC 10738) (41). The all-by-all comparison of genome similarity among the 42 *S. algae* assemblies revealed each isolate to represent a distinct strain, except isolates HUD-H4 and HUD-I2, which were retrieved from the same individual before and after antibiotic treatment (see Table S1 in the supplemental material). Higher JSI values were also observed for the HUD-G3::20533 and 5043::A97 isolate WGS assembly comparisons. Taken together, our analyses showcased substantial phylogenomic diversity within this sequenced group of isolates.

Next, we focused on analyzing the genomic diversity of our *S. algae* strains. To estimate the total gene pool of the 42 *S. algae* strains and assess their core and accessory gene repertoire, we performed a pangenome analysis. Upon successive addition of each genome, the number of protein-coding gene clusters containing homologs in the pangenome maintained a positive slope, indicating that the pangenome is essentially open, and reached a total of 10,122 (Fig. 2A). This indicated a substantial genetic diversity among the genomes of the sequenced *S. algae* strains, consistent with their diverse origins in terms of time, host, disease, and geography (Table S1), as well as their phylogenomic distance (Fig. 1). Conversely, the number of conserved genes decreased with the addition of each genome, resulting in 3,427 core genome genes (Fig. 2A). Thus, the *S. algae* pangenome is far from saturated. Overall, core genes (i.e., genes shared by all isolates) represented 37.65% of the pangenome, while shell genes (genes shared by 2 or more isolates) and cloud genes (genes unique to a single isolate) comprising the accessory genome represented 28.56% and 33.79%, respectively (Fig. 2B). Evaluating pangenome homologous gene cluster frequency demonstrated that genomic presence for the majority of accessory genome content was sparse, with more than half of the clusters being present in 25% or less of the *S. algae* isolates (Fig. 2C). Remarkably, these intra-specific proportions are similar to those reported for *Shewanella* in other comparative genomic studies involving a lower number of genomes belonging to diverse species (13, 14). Detailed content from our pangenome analysis is provided in Data Set S1 in the supplemental material.

To contextualize the relationships between isolates based on their pangenome, we clustered the strains based on their accessory genome content (Fig. 2E). Clades formed when comparing homologous gene cluster profiles were similar to those observed when assessing relationships with JSI (Fig. 1). While isolates HUD-H4 and HUD-I2 shared the same distinct accessory genome profile, this clustering also revealed substantial similarities in accessory genome content between the clinical isolate pairs HUD-G3::20533 and 5043::A97, consistent with their WGS-based similarities (Fig. 1).

Next, to gain functional insights into the core, shell, and cloud genome contents of the 42 *S. algae* strains, KEGG pathway annotations were generated for each pangenome category (Fig. 2F). Most core genome genes belonged to amino acid and carbohydrate metabolism, representing together about 25% of the core genome content. Other significantly populated KEGG pathways in the core genome were energy metabolism, metabolism of cofactors and vitamins, and signal transduction, altogether representing about 20% of the core genome content. Carbohydrate metabolism pathways also comprised the majority of accessory (cloud and shell) genes. Notably, the second most populated accessory genome category was glycan biosynthesis and metabolism, suggesting a rather diverse glycobiology among *S. algae* strains that so far has only been incipiently explored (44). Signal transduction, secondary metabolite biosynthesis, and cellular community pathways were also significantly represented in the accessory genome.

Among cloud genes, we identified some noteworthy features, including antibiotic resistance determinants such as *sul2*, *aph(3'')-Ib*, *aph(6)-Id*, *floR*, *tetR*, and *tetD* (strain 150735, locus tags JKK46_16645, JKK46_16650, JKK46_16655, JKK46_16670, JKK46_16685,

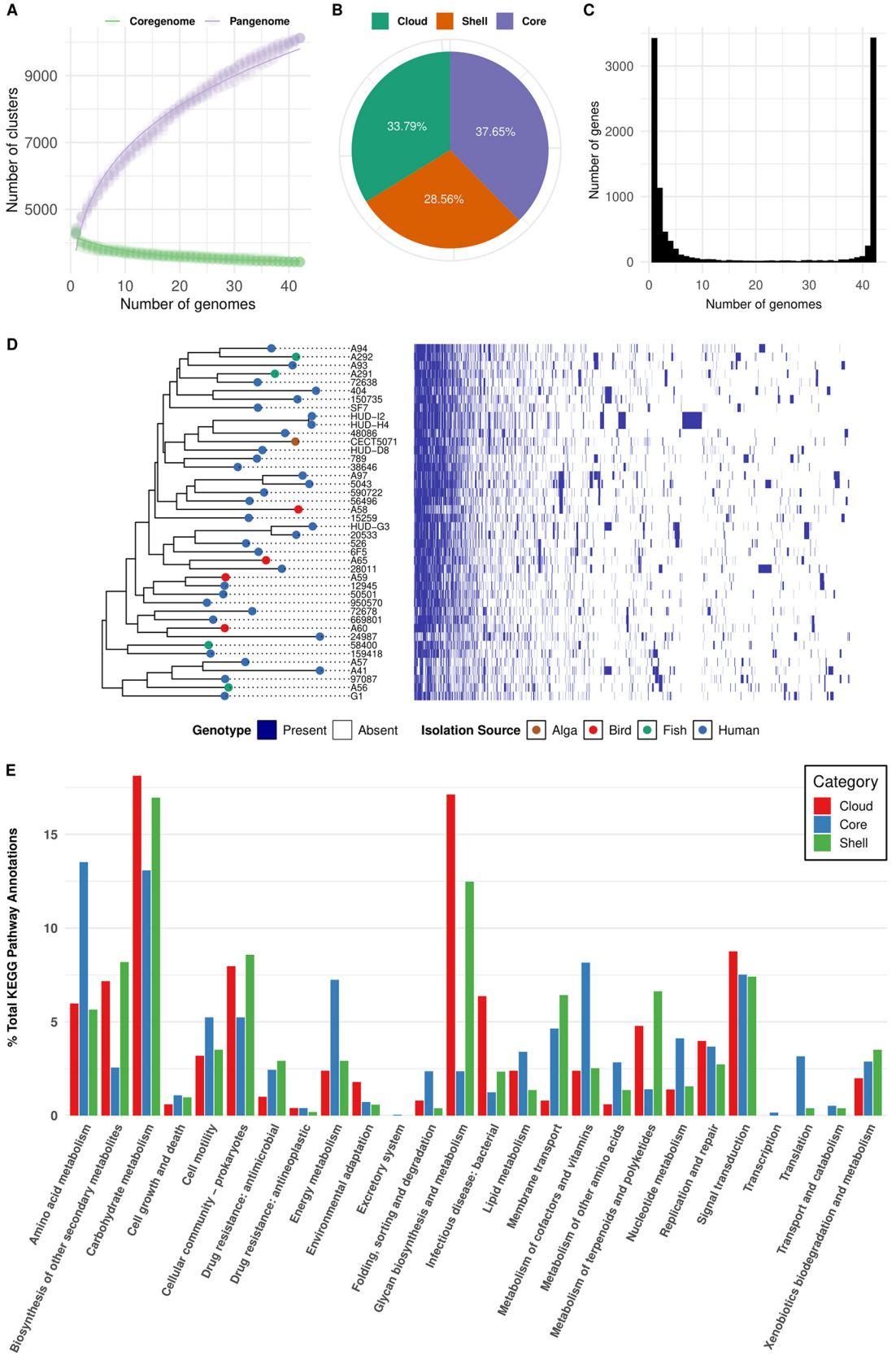

FIG 2 Pangenome analysis of *S. algae* strains. (A) Pangenome rarefaction curves illustrating the growth in pangenome size (top, purple) and stabilization of the core genome (bottom, green) after subsequent genome additions during construction.

(Continued on next page)

and JKK46_16690). An additional nitrite reductase, besides the two constitutive *nrfA* paralogs, was identified in strain 97087 (locus tag I6M44_21660), which also contained a copper resistance operon (locus tags I6M44_04940, I6M44_04945, I6M44_04950, and I6M44_04955). Accessory genes included genes associated with representative toxin-antitoxin systems (e.g., *higBA*, strain CCUG 12945, locus tags I6M59_21290 and I6M59_21295), restriction-modification systems (e.g., CCUG 20533, locus tags I6M54_19235 to I6M54_19245), and a fimbrial biogenesis cluster (strain CCUG 24987, locus tags I6M53_13090 to I6M53_13105). A urease operon found in strain CCUG 789 (locus tags I6M58_18955 to I6M58_18985) is exceptional, as *S. algae* has been reported to be urease negative (41). Besides, numerous putative c-di-GMP turnover genes and chemotaxis protein-coding genes were identified in the accessory pangenome (see below). Of note, the accessory nitrite reductase NrfA of *S. algae* 97087, a strain that lacks the dimethyl sulfoxide reductase operon (15), is more closely related to NrfA proteins of *Shewanella xiamenensis* and *Shewanella putrefaciens* than to either of the two NrfA paralogs seen in other strains of *S. algae* (see Fig. S1A in the supplemental material). This finding supports the possibility that horizontal gene transfer of anaerobic respiration pathways between *S. algae* and other *Shewanella* spp. has occurred (15). The genomic context for *nrfA-3* in *S. algae* 97087 suggests acquisition via transposition (Fig. S1B). Also of note, while all strains contained *qnrA* and *bla*-OXA-55 genes, the human wound isolate 150735 carried additional virulence factors, including antibiotic resistance determinants encoded by a chromosomal genomic island at positions 3670095 to 3775315 that contained IS*91* and Tn*3* family transposase elements and a type IV conjugal transfer system, as well as phage-related proteins (data not shown).

Collectively, the components of our pangenome analysis highlight the distinct genomic features and rich genetic diversity among these 42 isolates. As part of our ongoing efforts to understand *S. algae* biofilm physiology in response to environmental stimuli, the following sections showcase a detailed analysis of cyclic di-GMP metabolism and chemosensory networks.

The *Shewanella algae* c-di-GMP proteome is mostly conserved among strains. Biofilm formation in *Shewanella* spp. is typically controlled by the bacterial second messenger c-di-GMP (15, 45, 46). To explore the diversity of c-di-GMP metabolizing proteins in the 42 *S. algae* strains, we first identified putative c-di-GMP turnover proteins in the type strain CECT 5071. This search revealed 63 open reading frames (ORFs): 33 proteins with a GGDEF domain (PF00990 domain in Pfam [47]) being putative DGCs, 4 proteins with an EAL domain (PF00563 in Pfam) with putative c-di-GMP PDE activity, 19 hybrid proteins with a GGDEF-EAL domain combination, and 7 putative c-di-GMP PDEs with an HD-GYP domain. The type strain c-di-GMP turnover protein density was 12.79 per Mbp, which is close to the average for *Shewanella* spp., but much higher than the bacterial average of 4.11 per Mbp (calculated from the data in https://www.ncbi.nlm.nih.gov/Complete_Genomes/c-di-GMP.html).

With *S. algae* CECT 5071 being the type strain and the reference model, we thoroughly dissected its c-di-GMP turnover proteome. We performed sequence alignments of GGDEF, EAL, and HD-GYP domains, encoded in the genome of this strain, investigated the presence of key structural motifs, and reconstructed their phylogenies with respect to reference proteins. Of the 33 GGDEF domain proteins of *S. algae* CECT 5071, 27 domains contain the GG(D/E)EF signature motifs indicative of catalytic activity (see Fig. S2A in the supplemental material). Four domains from proteins WT_01800, WT_01911, WT_02417, and WT_02823 contain

FIG 2 Legend (Continued)
(B) Pie chart showcasing relative proportions of the core, shell, and cloud pangenome categories. (C) Gene frequency bar plot reflecting the number of homologous gene clusters present (*y* axis) in respective proportions of the isolate population used to construct the *S. algae* pangenome (*x* axis). Bars to the far right indicate gene clusters comprising the core genome. (D) Clustered pangenome dendrogram based on the respective accessory genome (gene presence/absence for shell and cloud genomic categories) profiles of each independent *S. algae* genome. Each column represents a homologous gene cluster with blue indicating presence and white indicating absence. Dendrogram tip point colors designate the isolation source. (E) Summary of *S. algae* pangenome KEGG pathway annotations. Bar colors correspond to the respective pangenome category (red, cloud; blue, core; green, shell), and bar heights indicate the percentage of total gene clusters comprising each category that received KEGG pathway annotations.

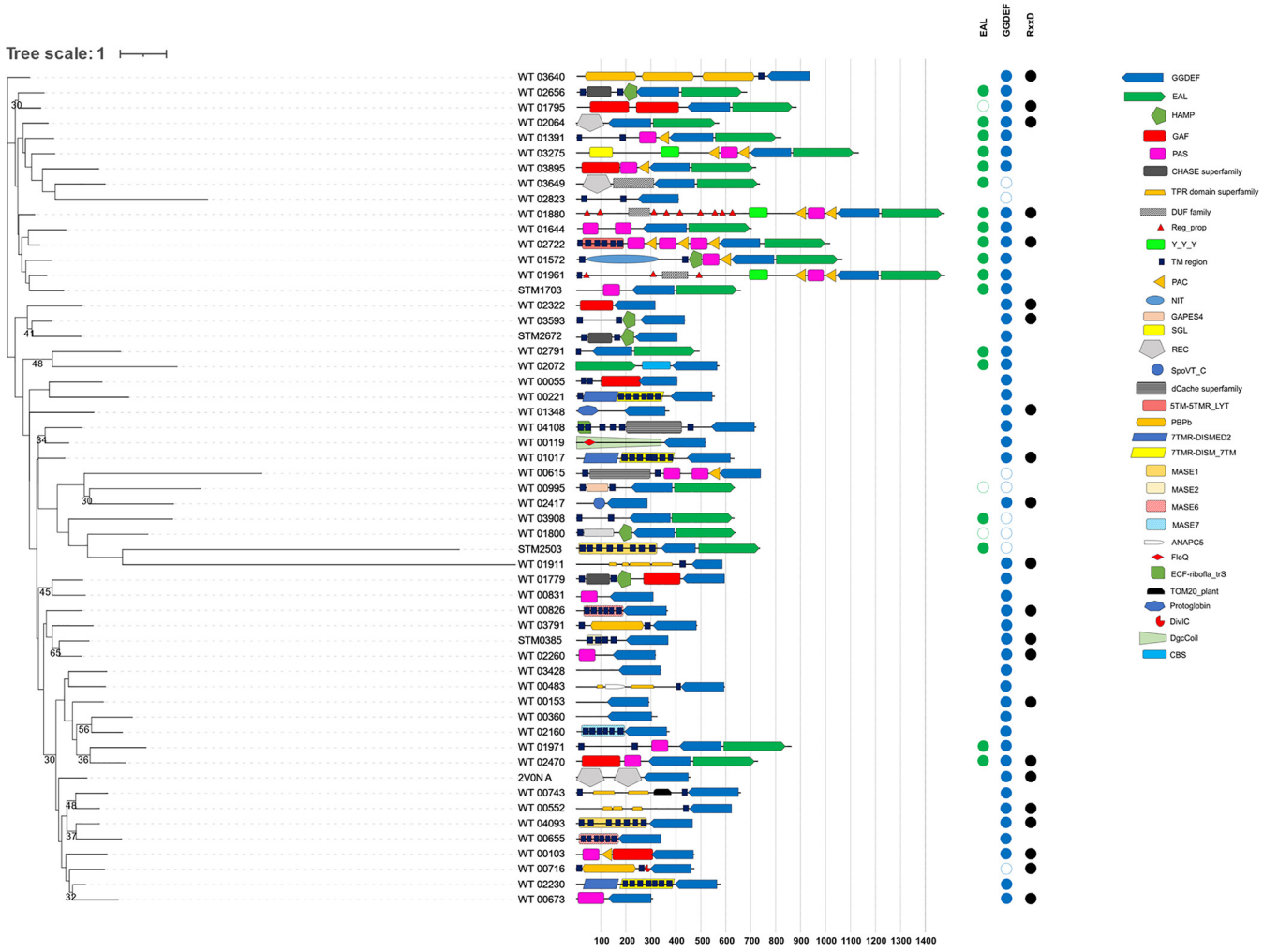

FIG 3 Phylogenetic tree of full-length GGDEF domains of *S. algae* CECT 5071 and domain architectures of the corresponding proteins. Experimentally characterized GGDEF domain proteins AdrA (STM0385), YfiN (STM2672), YciR (STM1703), and YfgF (STM2503) from *Salmonella* Typhimurium and PleD (PDB ID 2V0N) from *Caulobacter vibrioides* were chosen as reference points. The circles on the right indicate GGDEF domains with (filled blue circles) or without (open blue circles) the active-site GG(D/E)EF motif and the presence of the autoinhibitory RxxD motif (black circles) and/or the EAL domain with (filled green circles) or without (open green circles) the eponymous EAL motif. Node support values above 30% are indicated. Protein names are followed by their domain architectures, determined by searches with SMART, ScanProsite, HHPred, and CDVist. Domains identified by HHSearch as implemented in CDVist are displayed if the probability was ≥90.0%. Protein lengths (in amino acid residues) are indicated at the bottom. The domain names and their Pfam database (47) entries are as follows: GGDEF, PF00990; EAL, PF00563; HAMP, PF00672; GAF, PF01590; PAS (or PAS+PAC), PF00989; CHASE, PF03924; TPR, PF00515; DUF, PF11849, Reg_prop, PF07494; YYY, PF07495; NIT, PF08376; GAPES4, PF17157; SGL, PF08450; REC, Response_reg, PF00072; SpoVT_C, PF15714; dCache, PF02743; 5TM-5TMR_LYT, PF07694; PBPb, SBP_bac_3, PF00497; 7TMR-DISMED2, PF07696; 7TMR-DISM_7TM, PF07695; MASE1, PF05231; MASE2, PF05230; MASE3, PF17159; ANAPC5. PF12862; FleQ, PF06490; ECF-ribofla_trS, PF07155; TOM20_plant, PF06552; Protoglobin, PF11563; DivIC, PF04977. For MASE6 and MASE7, see Fig. S5; for DgcCoil, see Fig. S6.

neither the GG(D/E)EF motif nor additional consensus motifs required for catalytic activity, making their functionality as c-di-GMP synthases unlikely. Two proteins were excluded from this comparison. Protein WT_04201 is 169 amino acids long and consists solely of a stand-alone GGDEF domain with no N-terminal signaling domain and the eponymous active-site motif changed to NHHLF. WT_00162 is highly truncated, consisting of only 76 amino acids corresponding to the C terminus of the GGDEF domain. While GGDEF domains with altered signature motifs are usually catalytically inactive, they may play regulatory roles (16). An RxxD autoinhibitory I-site (48) involved in allosteric c-di-GMP binding that restricts the DGC activity was identified in 21 of the 31 full-length GGDEF domain proteins (Fig. S2A).

A phylogenetic reconstruction of GGDEF domains of *S. algae* CECT 5071 with respect to the reference GGDEF domains from other bacterial species is presented in Fig. 3. Thirty-one of the 33 GGDEF-domain proteins contain N-terminal fragments that could serve as potential sensor domains. Twenty-one of these proteins are predicted

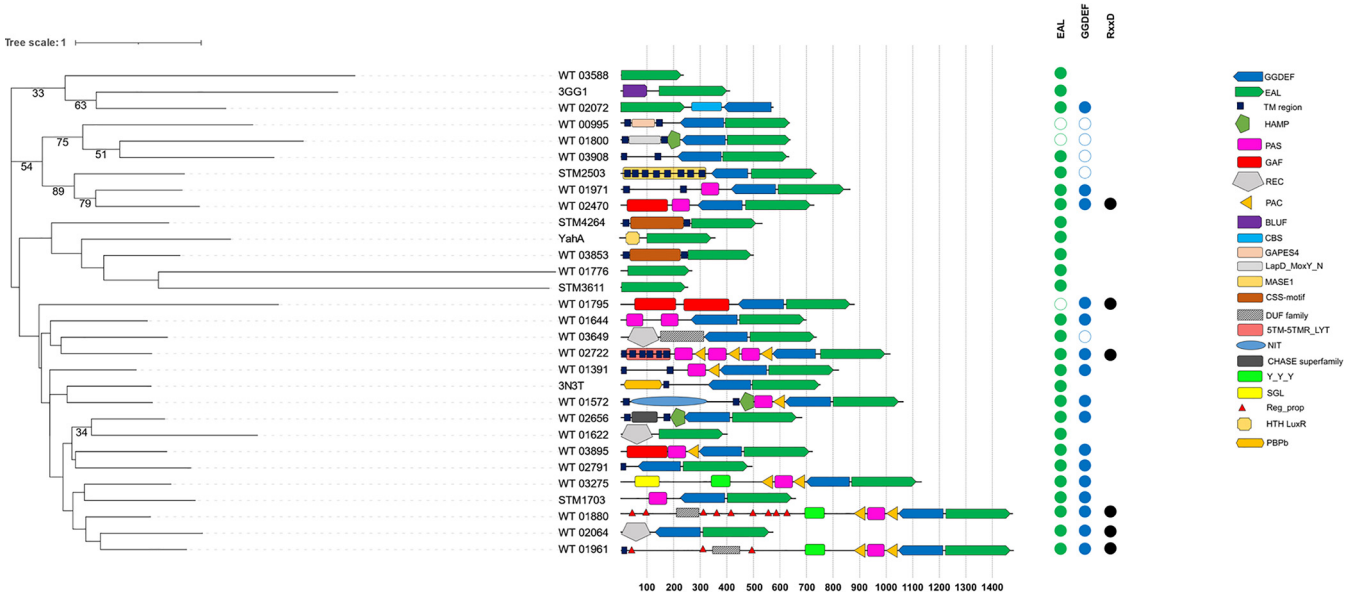

FIG 4 Phylogenetic tree of the EAL domains of *S. algae* CECT 5071 and domain architectures of the corresponding proteins. EAL domains of experimentally characterized proteins YciR (STM1703), YfgF (STM2503), YhjH (STM3611), and YjcC (STM4264) from *Salmonella* Typhimurium, YahA from *E. coli*, Tbd1265 from *Thiobacillus denitrificans* (PDB ID 3N3T), and BlrP1 from *Klebsiella pneumoniae* (PDB ID 3GG1) were chosen as reference points. Other details and domain shapes are as in Fig. 2, except for the BLUF (PF04940), CBS (PF00571), LapD_MoxY_N (PF16448), CSS-motif (PF12792), and DUF3369 (PG11849) domains.

to be membrane bound and contain one or more transmembrane helices with periplasmic or membrane-embedded N-terminal domains, most of which were not identified by SMART, Pfam, or CD-search. N-terminal sensor domains include the small ligand binding PAS/PAC domain and GAF domain in three proteins each. These domains are most commonly associated with GGDEF and EAL signaling domains, can bind small molecules such as NO, oxygen, and nucleotides, and sense light, temperature, and the redox status. Periplasmic substrate-binding domains (PBPb; PF00497 in Pfam) were found in WT_00716, WT_03640, and WT_03791, 7TMR-DISMED2 integral membrane domains in WT_00221, WT_01017, and WT_02230, and a dCache domain in WT_04108. The identification of novel N-terminal domains in these and other c-di-GMP turnover proteins is described in the next section.

The four EAL domain proteins retain all recognized active-site residues for c-di-GMP-hydrolyzing PDE activity (Fig. S2B). A phylogenetic reconstruction of *S. algae* CECT 5071 EAL domains with respect to reference EAL domains from other bacterial species is presented in Fig. 4. WT_01776 and WT_03588 represent stand-alone EAL domains. The putative c-di-GMP PDE WT_01622 is a response regulator of the PvrR/RocR family that has the REC-EAL domain architecture, combining the EAL domain with the phosphoacceptor (receiver [REC]) domain of two-component signal transduction systems (49) (Fig. 4). A novel divergent CSS motif domain (described below) was identified in WT_03853.

Of the 19 GGDEF-EAL domain proteins, 15 GGDEF domains contain a GGDEF motif, with 5 of them containing an RxxD I-site motif, while the conservation of the consensus motif indicates functionality for 16 EAL domains. The GGDEF domains of proteins WT_00615, WT_03649, and WT_03908 do not contain a GG(D/E)EF motif, but the EAL domains of these proteins could be functional. Two proteins, WT_00995 (ortholog of *E. coli* CsrD [50, 51]) and WT_01800 (ortholog of *Pseudomonas fluorescens* LapD [52, 53]), possess unconventional EVF and ELF active-site motifs, respectively. Based on the conservation of the catalytically important motifs, WT_00995 might possess PDE activity, as has been observed for PigX of *Serratia* sp. (54), while WT_01800 is likely not catalytically active. The EAL domain of WT_01795 has the eponymous EAL motif replaced by the DVR motif and also lacks other motifs involved in metal binding and catalysis,

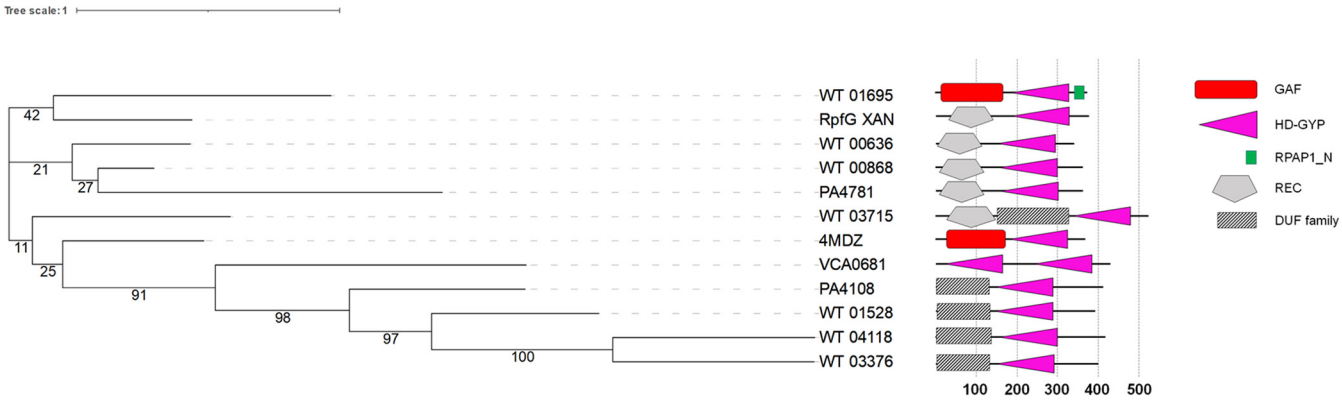

FIG 5 Phylogenetic tree of HD-GYP domains of *S. algae* CECT 5071 and domain architectures of the corresponding proteins. The experimentally characterized HD-GYP domain proteins from *Vibrio cholerae* (VCA0681, PDB ID 5Z7C), *Pseudomonas aeruginosa* PAO1 (PA4108 and PA4781; PDB ID 4R8Z), *Persephonella marina* (PDB ID 4MDZ), and *Xanthomonas campestris* (RpfG) were chosen as reference points for HD-GYP domains. Other details and domain shapes are as in Fig. 2, except for the HD-GYP domain (COG2206 in the COG database [72]) and DUF3391 (PF11871 in Pfam).

suggesting a lack of catalytic activity (Fig. 3; Fig. S2A and B). With respect to the N-terminal signaling domains, 10 proteins contain at least one PAS domain and 3 proteins contain at least one GAF domain (Fig. 3). Two proteins contain a REC (receiver) domain characteristic of two-component phosphotransfer systems, and 11 proteins contain at least one transmembrane helix (Fig. 3).

The GGDEF and EAL domain superfamilies are two of the most abundant bacterial protein superfamilies. Analysis of the phylogenetic relationships of 50 of the 52 GGDEF domains, excluding WT_00162 and WT_04201, with a truncated and a highly divergent GGDEF domain, respectively, showed that GGDEF domains of *S. algae* are distantly related, clustering rather with GGDEF domains of similar domain architecture from other species (Fig. 3; Fig. S2A) as previously observed (55, 56). Similarly, *S. algae* EAL domains are distantly related, with the closest homologs of similar domain architecture found in other species (Fig. 4). These results generalize previous findings of EAL domains to cluster according to their domain architectures rather than according to species (55), suggesting that most of the respective genes were vertically inherited from a common ancestor and subsequently evolved with respect to domain structure and sequence diversification, although horizontal gene transfer also played a role (7).

The HD-GYP domain (57) is characterized by c-di-GMP (and occasionally cGAMP) specific PDE activity (58, 59). Seven HD-GYP domain proteins with characteristic consensus amino acid signature motifs suggesting catalytic activity were identified. The HD-GYP domains of functionally or structurally characterized reference proteins of the two HD-GYP subfamilies and the cGAMP PDE HD-GYP domain of VCA0681 (60) were aligned with *S. algae* HD-GYP domains (Fig. S2C). Three *S. algae* HD-GYP domain proteins have an N-terminal REC domain, three have the DUF3391 (Pfam PF11871) domain, which has not yet been experimentally characterized but whose structural model is available on the Pfam website (61), and one protein has a GAF sensor domain. Phylogenetic analysis (Fig. 5) revealed that HD-GYP domain proteins in *S. algae* CECT 5071 belong to the two previously identified subfamilies (58). Based on the conservation of the Fe-coordinating Glu residue (E185 in the structure of *Persephonella marina* PmGH (PDB ID 4MDZ) (Fig. S2C), four HD-GYP domains of *S. algae* (WT_00636, WT_00868, WT_01695, and WT_03715) contain a predicted trinuclear metal center and are likely to hydrolyze c-di-GMP to GMP, whereas the other three HD-GYP domain proteins (WT_01528, WT_03378, and WT_04118) contain a binuclear center and only hydrolyze c-di-GMP to the linear dinucleotide pGpG (61). No close homolog of the distinct cGAMP-specific HD-GYP domain of the *Vibrio cholerae* protein VCA0681 (60) is encoded in *S. algae* CECT 5071.

We subsequently retrieved all putative c-di-GMP metabolizing proteins from other 41 *S. algae* genomes. The number of c-di-GMP metabolizing proteins identified ranged from 61 (*S. algae* A57) to 67 (*S. algae* A41), with relative c-di-GMP turnover protein

densities between 12.38 per Mbp (*S. algae* A56 and *S. algae* A97) and 13.53 per Mbp (*S. algae* A41) (Fig. 6A). The distribution of conserved, nonconserved, and distantly related c-di-GMP turnover orthologs of the type strain in the other 41 isolates is presented in Fig. 6B. Two proteins, WT_00103 and WT_0162, are absent in most or all other strains. Thus, WT_00103 orthologs were only found in 11 of the 41 strains, suggesting that this putative DGC is uncommon. WT_00103 is located in the *relA-recN* genomic region at positions 3302276 to 3319231 in the genome of strain CECT 5071 (39), and part of this region is contained in a genomic island (see Fig. S3A in the supplemental material). In *S. algae* strain G1, which lacks this DGC, the same genomic region is divergently transcribed and located at a different position in the complete genome sequence (Fig. S3B). The truncated, stand-alone GGDEF domain protein WT_00162 is unique to the type strain and might be vestigial. Divergent orthologs of the GGDEF domain proteins WT_02322 and WT_02823 and the GGDEF-EAL domain protein WT_03275 were frequently found in other strains, suggesting that these proteins might be under evolutionary pressure in *S. algae*.

The phylogenetic positions of GGDEF, EAL, and HD-GYP domains identified in the 41 other *S. algae* strains with reference to *S. algae* CECT 5071 proteins are shown in Fig. 7A to C, and the domain architectures and key features of representative proteins are presented in Fig. S4 in the supplemental material. These include the membrane-bound GGDEF domain proteins 159418_02667 and 97087_01449 (5TM helices each) and 789_00791 (7TM helices), with only 97087_01449 showing the RxxD inhibitory I-site. A cytoplasmic PAC-GGDEF domain protein was identified in strains A41 and G1 (shown is that of strain G1, G1_04334), in a distinct genomic island that also contained a cytoplasmic GGDEF-EAL domain protein with 3 PAS domains and one GAF domain (G1_04336), along with chemosensory proteins (Fig. S3B). A membrane-bound GGDEF-EAL domain protein with a HAMP domain typical of bacterial transmembrane sensory proteins was found in strain A291 (A291_04011) lacking GGDEF and EAL motifs and an I-site, suggesting catalytic inactivity. Unique cytoplasmic EAL domain proteins were identified in strains of different genetic backgrounds, including HUD-G3_03949, which shows a similar architecture to WT_01622, including an N-terminal REC domain. A putative HD-GYP c-di-GMP PDE, 50501_04362, was found to be unique to the clinical isolate CCUG 50501. This PDE is predicted to be cytoplasmic, with an N-terminal GAF domain that participates in small molecule binding and protein dimerization (62), and a C-terminal CBS domain involved in the regulation of functions in response to the cellular energy status (63). This unique HD-GYP protein is located upstream of a helix-turn-helix (HTH)-containing transcriptional regulator and an HBM-LBD-type chemoreceptor (50501_04360). The three-gene cluster is flanked by transposases (locus tags I6M49_16805 to I6M49_16835 in the draft genome sequence; GenBank accession no. JADZHO000000000.1), a tell-tale sign of acquisition via horizontal gene transfer. Two proteins, 72678_00888 and HUD-G3_02699, are present in most of the strains (Fig. 6). Taken together, the c-di-GMP proteome of *S. algae* is conserved to a large extent, but exhibits some degree of variability across strains with clear evidence of horizontal acquisition of c-di-GMP turnover genes in some isolates. Cyclic di-GMP gene gain, loss, or sequence divergence might relate to adaptation to specific environmental conditions that are yet to be determined.

The *S. algae* c-di-GMP proteome encodes novel N-terminal sensor domains and a new CSS domain variant. Examination of the amino acid sequences of the c-di-GMP turnover proteins from *S. algae* revealed that a substantial fraction of them had uncharacterized N-terminal sequences that could represent new sensor domains. As a first step toward the characterization of the sensory capacity of *S. algae*, we constructed the alignments of these homologous domains from the respective proteins found in various organisms, predicted their structural organization and transmembrane topology, identified the conserved residues, and submitted them to the Pfam protein domain database (47). We provide here the description of five novel membrane-embedded domains, an N-terminal cytoplasmic coiled-coil domain, and a distinct variant of the previously characterized CSS motif domain.

Previous studies resulted in the identification of a variety of periplasmic sensor

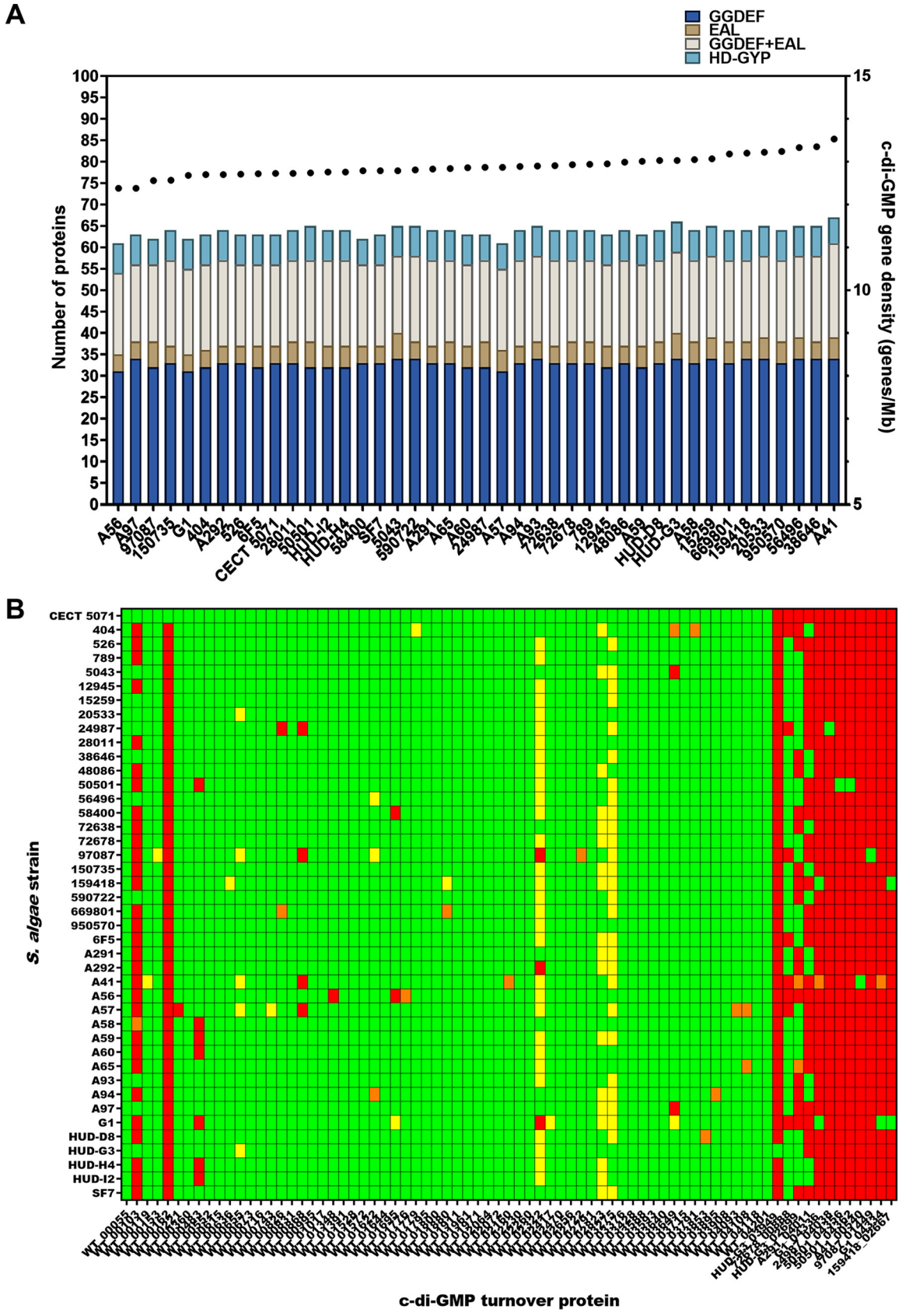

FIG 6 Cyclic di-GMP turnover proteome of *Shewanella algae* strains. (A) Distribution of GGDEF, EAL, GDDEF+EAL, and HD-GYP domains in *S. algae* genomes (bars, left axis) and total putative c-di-GMP turnover protein density per genome (dots, right axis). (B) Heat map showing

(Continued on next page)

domains (34, 64) and a relatively small number of integral membrane domains, which included the NO-binding MHYT domain, named after its conserved sequence motif (65, 66), 5TMR of the 5TMR-LYT and 7TM-DISM_7TM domains (67), and five MASE (membrane-associated sensor) domains (18, 68, 69), of which only MASE1 (Pfam domain PF05231) has been experimentally characterized and was found to sense aspartate in *E. coli* PdeF (YfgF) and control the proteolytic turnover of DgcE (70, 71). Analysis of the *S. algae* CECT 5071 GGDEF domain proteins WT_00655 (GenBank accession no. QTE77313.1), WT_00826 (GenBank accession no. QTE77466.1), and WT_02160 (GenBank accession no. QTE78688.1) revealed that their GGDEF domains are preceded by hydrophobic fragments that consist of six predicted transmembrane (TM) segments. Iterative PSI-BLAST searches revealed that the TM regions of WT_00655 and WT_00826 are related and contain the same 6TM sensor domain (hereafter MASE6), whereas WT_02160 contains another 6TM domain, hereafter MASE7 (see Fig. S5A to C in the supplemental material). In addition to the GGDEF domains, MASE6 and MASE7 were found in association with histidine sensor kinases, and MASE7 was also associated with the adenylate cyclase output domain (PF00211 in Pfam), confirming that they both comprise promiscuous sensor domains in diverse bacterial receptors. Analysis of the membrane topology of these domains revealed that they contain short N-terminal hydrophilic regions located in the cytoplasm, which are followed by six TM segments that are connected by very short cytoplasmic and periplasmic loops. These loops are too short to form any separate ligand-binding domains in the periplasm and show poor sequence conservation, which suggests that signals are likely sensed by the membrane-embedded portions of MASE6 and MASE7 domains. Indeed, both domains contain several conserved Phe, Tyr, and Trp residues that could be involved in binding aromatic compounds. Alternatively, these domains could modulate the catalytic activity of the downstream domains through protein-protein interactions, as has been recently demonstrated for the MASE1-containing DgcE (71). Comparative analysis of *S. algae* genomes revealed one more 6TM integral membrane sensor domain in a DGC from strain 97087, named MASE8, which is widespread among various classes of *Proteobacteria* and is also found in representatives of some other phyla (Fig. S5D). In addition to DGCs, MASE8 was found in association with adenylate/guanylate cyclase output domain and in histidine sensor kinases (Fig. S5D), which confirmed its identity as a sensor domain.

One more integral membrane sensory domain, seen in DGCs from 32 out of 42 strains of *S. algae*, albeit not in the type strain, is distantly related to previously characterized proteins. A shorter, divergent stand-alone version of this domain, which corresponds to the *E. coli* protein YhhQ, a member of the COG1738 family in the COG database (72), has been described as an essential component of the transport system for queuosine precursor ($preQ_0$) 7-cyanodeazaguanine (73, 74). The YhhQ protein has 6 predicted TM segments and is listed in Pfam as "Putative vitamin uptake transporter Vut_1" (PF02592). In *S. algae* and other gammaproteobacteria, the YhhQ-related N-terminal sensor domain has an extra TM segment and combines with the GGDEF domain to form a widely conserved DGC (see Fig. S6A in the supplemental material). This domain, which we named VUPS (vitamin uptake-like sensor), is found in a variety of bacteria in combination with PAS, GGDEF, and EAL domains, as well as with histidine sensor kinases. Several archaea also encode histidine sensor kinases with VUPS as their sensor domain (Fig. S6A).

The DGC WT_00119 (GenBank accession no. QTE76827.1) of *S. algae* CECT 5071 combines the GGDEF domain with a 350-amino-acid long N-terminal α-helical cytoplasmic domain. Analysis by programs such as PCOILS, MultiCoil2, and DeepCoil (75–77) indicated that a large part of this domain consists of a coiled-coil structure

FIG 6 Legend (Continued)
the degree of conservation of putative orthologs of CECT 5071 c-di-GMP turnover proteins in other 41 *S. algae* genomes. Green, proteins that are ≥90% identical in length ($P_{align} \geq 90\%$) and have ≥95% amino acid sequence identity ($P_{ident} \geq 95\%$); yellow, proteins with either $P_{align} > 90\%$ and $P_{ident} > 75\%$ or $P_{align} > 50\%$ and $P_{ident} > 90\%$; red, proteins with $P_{align} < 50\%$ and $P_{ident} < 75\%$; orange, annotated proteins showing single nucleotide changes disrupting the open reading frames or split across contigs.

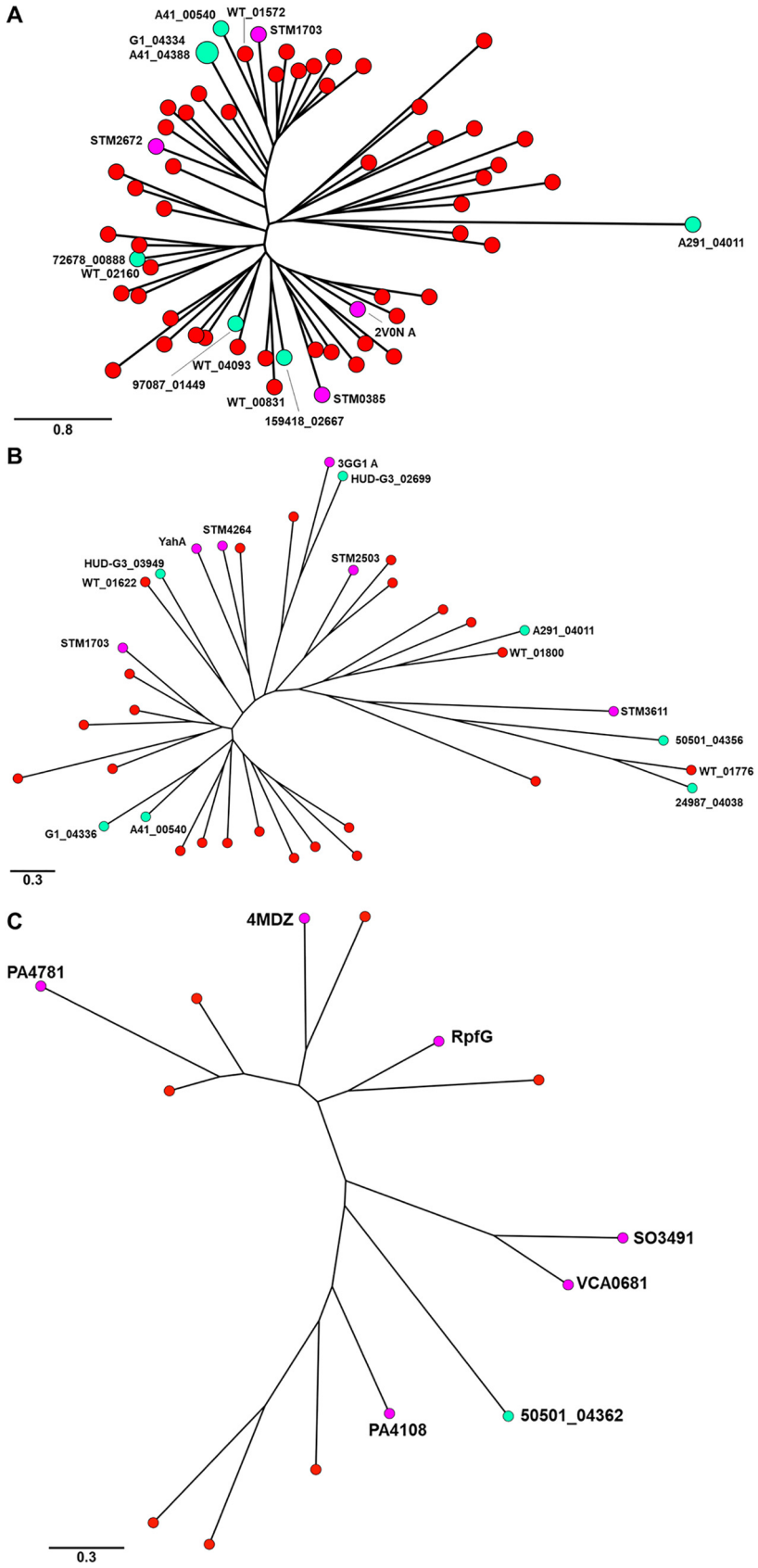

FIG 7 Phylogenies of the GGDEF, EAL, and HD-GYP domains of *S. algae* c-di-GMP turnover proteins not encoded by the type strain. The phylogenies of the GGDEF (A), EAL (B), and HD-GYP (C) domains identified

(Continued on next page)

(Fig. S6B). Furthermore, the MARCOIL/LOGICOIL combination (78) predicted the quaternary structure of this domain as either a tetramer (four-helix bundle [79]) or an antiparallel dimer. The combination of this domain with the GGDEF domain is widespread in gammaproteobacteria and two proteins with such a domain architecture have been experimentally characterized as active DGCs in *V. cholerae* (CdgG, locus tag VC0900; GenBank accession no. AAF94062) and *P. aeruginosa* (DgcH, locus tag PA5487; GenBank accession no. AAG08872) (80, 81). Accordingly, we designated this domain DGCcoil (DGC-associated coil). In contrast to the periplasmic α-helical sensor domains containing a predicted coiled coil, such as HBM (PF16591), DAHL (PF19443), and the sensor domain of TorS (PDB ID 3I9W), which have been previously described in DGCs, chemoreceptors, and histidine sensor kinases (82–85), the DGCcoil domain is cytoplasmic and shows no sequence (or predicted structural) similarity to either of these domains. It remains to be seen whether it binds any ligand, is responsible for some kind of protein-protein interaction, or just stimulates dimerization of the associated GGDEF domains.

Finally, *S. algae* DGC WT_00055 (GenBank accession no. QTE76780.1) consists of the cytoplasmic GAF and GGDEF domains, which are preceded by a short N-terminal transmembrane region that consists of two TM segments. Sequence alignment of this 2TM hairpin (Fig. S6C), which is found exclusively in *Shewanella* and *Aeromonas* spp., shows few conserved residues, suggesting that this hairpin might serve simply as a membrane anchor, rather than a full-fledged sensor domain.

An interesting new domain was also found in *S. algae* c-di-GMP phosphodiesterases. Sequence analysis of the *S. algae* EAL domain protein WT_03853 (GenBank accession no. QTE76250.1) showed that it contains an unusual N-terminal periplasmic sensor domain. While this domain was not recognized by any Pfam domain model, it was clearly related to the CSS motif sensor (PF12792) domain (86), but with an additional conserved CxxC motif (hereafter, CSS_CxxC domain) (Fig. S6D).

Proteins with the CSS-EAL domain architecture, comprising COG4943 in the COG database (72), are widespread in bacteria, with multiple paralogous genes identified: for example, in *E. coli* (5 genes), *Salmonella enterica* (5 genes), and *Pseudomonas aeruginosa* (3 genes). Figure S6D shows an alignment of these diverse CSS domains with the experimentally studied CSS domains from PdeB and PdeC proteins of *E. coli* (86). A typical CSS motif sensor domain contains two conserved Cys residues that can form a disulfide bridge and undergo dithiol-disulfide transitions, allowing regulation of the PDE activity of the cytoplasmic EAL domain in response to the redox conditions in the periplasm (86). One of the *P. aeruginosa* proteins with the CSS-EAL domain architecture, PA2818, activates biofilm formation in response to subinhibitory concentrations of the aminoglycoside antibiotic tobramycin (87). *S. algae* encodes only one CSS domain-containing protein, WT_03853, but its CSS motif domain carries two additional Cys residues that form a conserved CxxC motif. We identified CSS_CxxC domains in other *Shewanella* spp. as well as other members of *Altermonadales* (*Aeromonas*, *Oceanimonas*, *Pseudoaltermonas*, and *Thalassomonas* spp.), *Vibrionales* (*Vibrio*, *Enterovibrio*, and *Photobacterium* spp.), and in some acidobacteria ("*Candidatus* Koribacter versatilis" and *Terracidiphilus* sp.). Representative sequences are shown in Fig. S6D.

The CxxC motif is characteristic of redox proteins that are involved in the formation, isomerization, or reduction of disulfide bonds (88, 89). The presence of two distinct pairs of cysteines that are capable of dithiol-disulfide transitions may allow such a CSS domain to perform a wider range of responses to redox changes or respond to different redox changes that ultimately lead to alterations in c-di-GMP levels. *E. coli* and sev-

FIG 7 Legend (Continued)
in the 41 *S. algae* genomes in the context of type strain domains are displayed by GrapeTree representations (146). Nodes representing CECT 5071 proteins are in red, reference c-di-GMP turnover proteins (listed in the legends to Fig. 1 and 3) are in magenta, and distinct c-di-GMP turnover proteins from other strains are in turquoise. *S. algae* proteins are shown under their Prokka gene tags (see Table S2 for GenBank accession numbers).

eral other organisms appear to accomplish this task by using several different CSS-EAL proteins. The full breadth of redox responsive c-di-GMP turnover proteins remains to be characterized.

In addition, *S. algae* DGC WT_02823 (GenBank accession no. QTE79282.1) and dual GGDEF-EAL domain proteins WT_01391 (GenBank accession no. QTE80346.1), WT_03908 (GenBank accession no. QTE76295.1), and WT_01971 (GenBank accession no. QTE78520.1) contain uncharacterized periplasmic N-terminal domains that are found in many *S. algae* strains and across *Shewanella* spp. but rarely, if ever, outside this genus.

In summary, initial characterization of c-di-GMP turnover proteins in *Shewanella algae*, a marine organism with a high c-di-GMP IQ, has unraveled a number of novel integral membrane, periplasmic and cytoplasmic N-terminal signaling domains whose sensory specificity and/or protein-protein interaction capacity remains to be determined.

Chemosensory pathways of *S. algae* are conserved across strains and include an F7 pathway of unknown function. Cyclic di-GMP homeostasis and biofilm formation are functionally intertwined with chemosensory pathways, including chemoreceptors. There is a significant number of studies showing that chemoreceptors (90–94) and chemosensory signaling genes (95–98) affect biofilm formation/dispersion and/or c-di-GMP levels. In *P. aeruginosa*, there is evidence for two chemosensory pathways, *wsp* and *chp*, that sense surfaces and modulate in turn c-di-GMP levels initiating biofilm formation (99). Current information suggests that there are multiple mechanisms by which chemosensory signaling modulates biofilm formation, and there is a considerable research need for the identification of the underlying molecular details. To this end, we present here an analysis of the repertoire of chemoreceptor and chemosensory signaling genes in *S. algae*.

S. algae CECT 5071 encodes 29 chemoreceptors that are likely to stimulate two chemosensory pathways conserved across all 42 strains (see Fig. S7 in the supplemental material). Classification of pathway-signaling genes by the MIST 3.0 database indicates the presence of the F6 pathway, likely to be involved in chemotaxis, as well as the F7 pathway. The gene arrangement of the latter pathway shows similarities to the *P. aeruginosa che2* pathway, which is of unknown function but not involved in chemotaxis (100). Chemotaxis experiments with *P. aeruginosa* in-frame deletion mutants with mutations in *cheB2*, *cheA2*, and *mcpB/aer2* (the only receptor predicted to stimulate the *che2* pathway) had chemotaxis and aerotaxis phenotypes indistinguishable from those of the wild type (101). The number of chemoreceptors in *S. algae* strains ranged from 28 to 31, which suggests a relatively conserved chemosensory repertoire, in contrast to the diversity found in *P. aeruginosa* or *Pseudomonas putida* strains. While genomes of some bacteria harbor an extremely large number of chemoreceptor genes, such as 90 for *Caryophanon latum* (102), the average in bacteria is 14 (103). The presence of 29 chemoreceptors thus reflects an above average chemosensory capacity for *S. algae*. So far there are no experimental data available on chemoeffectors recognized by any of these receptors. Of the 29 chemoreceptors of strain CECT 5071 (Fig. 8A), 12 present a ligand-binding domain (LBD) of the Cache superfamily that corresponds to the largest family of extracytoplasmic bacterial sensor domains (64). Thus, of these, 3 chemoreceptors contain a single Cache (sCache) domain of subtype 2, and 9 chemoreceptors contain a double Cache (dCache) domain (7 of subtype 1, 1 of subtype 3, and 1 dual Cache_3-Cache_2). Along with the four-helix-bundle (4HB) domain, the dCache domain is the prevalent LBD in prokaryotes and is homologous to, but distinct from the PAS superfamily, containing a long N-terminal α-helix and two α/β-type modules (64). Many dCache domains bind different amines (104), and most known chemoreceptors for amino acids possess a dCache domain (105). In other species, chemoreceptors with sCache_2 domains were found to primarily bind small organic acids (64, 105).

Two chemoreceptors of the type strain present a four-helix bundle (4HB) 4HB_MCP_1 domain, which is also widespread in *Bacteria* and *Archaea* and is known to bind a diverse array of ligands (79), such as amino acids (105), C_6 ring carboxylic acids (105), boric acid (106), tricarboxylic acid (TCA) cycle intermediates (107), or polycyclic

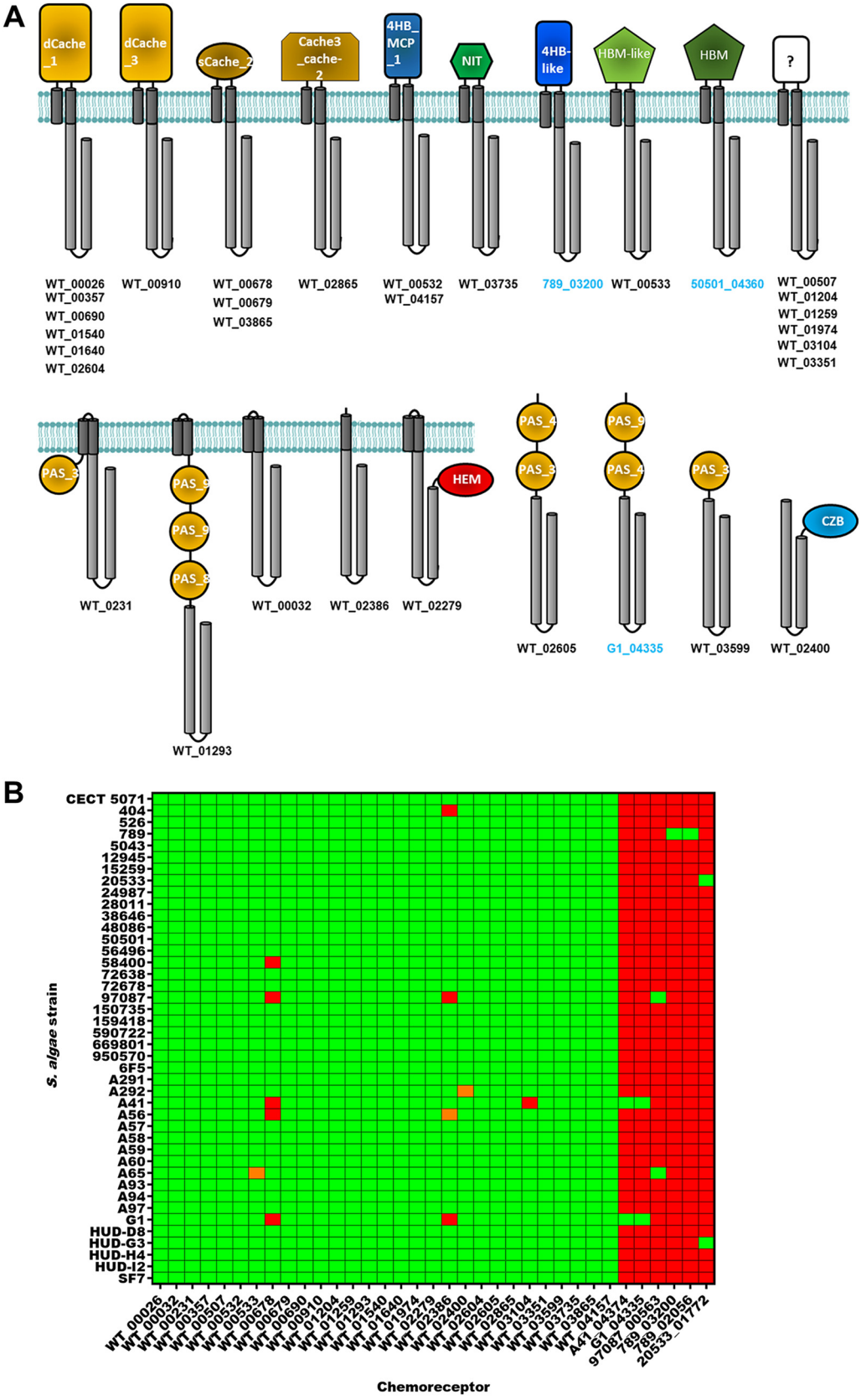

FIG 8 *Shewanella algae* chemoreceptors. (A) Domain architectures of *S. algae* chemoreceptors. Shown are the topologies of the reference chemoreceptor proteome of the type strain CECT 5071 (black lettering), as well as chemoreceptors identified

(Continued on next page)

aromatic hydrocarbons (108). A single chemoreceptor in the *S. algae* chromosome presents an NIT domain, predicted to bind nitrate and nitrite (109). By analogy to the sole so-far-characterized nitrate chemoreceptor McpN (110), this chemoreceptor was found to be downregulated in strain CECT 5071 in the presence of a high extracellular nitrate concentration (15), therefore supporting the notion of a role in nitrate recognition and nitrate-mediated signal transduction. Homology modeling identified an HBM-like LBD in WT_00533 (82) that was not identified by Pfam.

Six chemoreceptors of strain CECT 5071 exhibit unknown LBDs. Four MCPs, WT_01293, WT_02279, WT_00032, and WT_00231, exhibit a noncanonical topology characterized by two consecutive transmembrane regions without an LBD. WT_01293 protein has three PAS domains, WT_02279 has a C-terminal LBD, and WT_00231 has an N-terminal PAS domain. The latter receptor has the typical topology of an Aer receptor, which, by analogy to the studied homologs with the same topology, is likely to mediate aerotaxis (the PAS domain contains bound heme for oxygen sensing) (111).

WT_01293 has similarities to the Aer2/McpB chemoreceptor of *P. aeruginosa* (99, 112). The presence of multiple PAS domains is a characteristic feature observed for many Aer2/McpB homologs (38). By analogy to the P. aeruginosa chemoreceptor, WT_01293 is proposed to stimulate the F7 pathway (homologous to the che2 pathway in P. aeruginosa), which may have a function unrelated to chemotaxis. WT_02279 contains a C-terminal hemerythrin domain, which in other species was found to bind oxygen or nitric oxide (113, 114), suggesting that this receptor may also mediate aerotactic responses.

Three soluble MCPs, WT_02605, WT_03599, and WT_02400, complete the chemosensory repertoire of the type strain. WT_2605 and WT_03599 contain 2 N-terminal PAS domains and 1 N-terminal PAS domain, respectively, whereas WT_02400 contains at the C-terminus a CZB (chemoreceptor zinc-binding, PF13682) domain, which is commonly found in proteins involved in chemotaxis, c-di-GMP signaling, and nitrate/nitrite sensing (115). The characterized CZB domain-containing chemoreceptor TlpD of *Helicobacter pylori* was found to mediate repellent responses to oxidative stress (116), as well as chemoattraction to bleach (117), and similar responses may be mediated by WT_02400. Figure 8B summarizes the presence and degree of conservation of chemoreceptors in all 42 *S. algae* strains. Thus, the WT_02386 ortholog was absent in strains 404 and G1, and the WT_03104 ortholog was not present in strain A41. Distant orthologs of WT_00678 sharing 77 to 79% sequence identity with the type strain protein were identified in five *S. algae* strains, suggesting significant sequence variability in this specific chemoreceptor. Chemoreceptors found in other *S. algae* genomes that lack an ortholog in the type strain did not show distinct topologies and LBD types from the ones mentioned above (Fig. 8A). Homology modeling predicted a 4HB-like LBD (118) in the receptor 789_03200 of *S. algae* CCUG 789.

Chemoreceptors can be classified according to the length of their signaling domain (119). Since this domain forms a coiled-coil structure, the length is usually expressed in heptad repeats (H). Studies of *P. aeruginosa* have shown that the length of the signaling domain can be associated with the pathway it stimulates (120). For example, the 22 *P. aeruginosa* PAO1 chemoreceptors of the 40-heptad repeats (40H) family all stimulate the *che* pathway for chemotaxis, whereas the sole chemoreceptor of the 36H class (Aer2/McpB) was predicted to stimulate the *che*2 pathway. By analogy to this study, we propose that the 26 40H chemoreceptors of *S. algae* stimulate the chemotaxis pathway, whereas the sole 36H receptor (WT_01293) is proposed to stimulate the F7 pathway. This prediction thus coincides with the above-mentioned similarities in receptor topology between *P. aeruginosa* Aer2/McpB and WT_1293. While genes encoding receptors stimulating the F6 pathway are often distributed across the genome, a typical feature of

FIG 8 Legend (Continued)
in other sequenced *S. algae* strains with distinct topologies (blue lettering). Chemoreceptors that were not annotated in databases but identified through homology modeling are indicated with the suffix "-like". (B) Heat map showing conserved (green), divergent (yellow) or absent (red) *S. algae* CECT 5071 chemoreceptors in all other 41 *S. algae* genomes. Color criteria are as in Fig. 6.

F7 pathways is that the receptor that stimulates this pathway is encoded in a gene cluster together with the signaling proteins. This is the case of the 36H receptor WT_01293 and its homologs in other *S. algae* strains (Fig. S7). Besides, gene synteny of this cluster (*cheYAW-mcp*, *cheRB*) is remarkably similar to that of cluster II of *P. aeruginosa*, which codes for the F7 pathway (*cheYAW-mcp*, *cheRB*). Overall, the *S. algae* chemosensory system exhibits significant parallelism with that of *P. aeruginosa* PAO1, which encodes 22 40H receptors, two 24H receptors, one 36H receptor, and one 44H receptor, with the 36H receptor (Aer2/McpB) stimulating the F7 pathway (99). The 36H Aer2/McpB receptor of *P. aeruginosa* and its homolog in *V. cholerae* are known to sense oxygen at the PAS domain harboring the heme cofactor (121, 122). A similar function is therefore plausible for the *S. algae* 36H receptor. The function of the F7 pathway has not been identified but appears to be associated with virulence in *P. aeruginosa* (100, 123).

Of note, analysis of the acute enteritis isolates G1 and A41 of *S. algae* revealed the presence of a putative integrative and conjugative element containing a putative aerotaxis receptor, frequently associated with virulence (124), a CheV homolog, a cytosolic GGDEF-EAL domain-containing protein with one GAF and 3 PAS N-terminal sensor domains, a PAS domain-containing chemoreceptor with a topology similar to that of the *P. aeruginosa* BdlA (biofilm dispersion locus A) chemoreceptor (92), and a divergently transcribed GGDEF domain-containing protein (Fig. S3B). Such a genetic environment suggests a potential role of cyclic di-GMP signaling in the coordinated regulation of chemotaxis and virulence in *S. algae*.

Concluding remarks. Phylogenomic and pangenome analyses of 42 *S. algae* isolates revealed a relatively high degree of divergence, with most isolates being genetically distinct. Thus, *S. algae* shows an endemic population structure; however, the occurrence of closely related strains distinct in space and time suggests that clones with stable genetic repertoire exist in the population. The notable features of the *S. algae* accessory genome included horizontally acquired respiration-related genes, and antibiotic resistance determinants, as well as genes for urea hydrolysis enzymes and chemosensory and c-di-GMP turnover proteins.

With 63 c-di-GMP turnover proteins and 29 chemoreceptors in the reference (type) strain, *S. algae* is near the top in *Bacteria* in terms of signaling capacity per Mb of the genome sequence. The apparent redundancy of putative c-di-GMP turnover proteins likely reflects a highly flexible c-di-GMP network. Analysis of structural features of GGDEF, EAL, and HD-GYP domains of the type strain suggests that 27 out of 33 GGDEF-containing proteins, 16 of 19 dual GGDEF-EAL proteins, and all 7 HD-GYP domains retain the ability to synthesize or hydrolyze c-di-GMP. In contrast, 12 GGDEF domains and 3 EAL domains deviate from the respective canonical motifs that are indicative of catalytic activity. Ongoing efforts at the functional analysis of the complete reference c-di-GMP proteome are expected to shed light on the catalytic activities of these DGCs or PDEs. Phylogenetic analyses of *S. algae* GGDEF, EAL, and HD-GYP domains using heterologous reference proteins, in combination with protein topology and domain architecture analysis, show that *S. algae* c-di-GMP turnover proteins generally cluster according to their domain architectures rather than according to species, suggesting inheritance from a common gammaproteobacterial ancestor. Analysis of the domain topology of proteins involved in c-di-GMP metabolism (Fig. 3 and 5) reveals that the majority of sensor domains are located in the cytosol, indicative of a central role of cytosolic signals in modulating c-di-GMP levels. In contrast to GGDEF and EAL proteins, HD-GYP proteins have only two types of sensor domain, GAF and DUF3391, which may suggest a narrow range of signals modulating the activity of this protein family. The significant number of RpfG family response regulators with the REC-HD-GYP domain architecture (61) suggests an important input from sensor kinases.

The occurrence of c-di-GMP turnover proteins in the accessory genome of certain isolates raises questions about their contribution to *S. algae* fitness inside or outside a host and the rewiring of the signaling network upon their introduction. The comparative analysis of the c-di-GMP turnover proteome of *S. algae* isolates also revealed that

the core proteins WT_02322, WT_02823, and WT_03275, while conserved in domain architecture, exhibit the highest degree of sequence variability across isolates, suggesting evolutionary pressure toward positive selection of those signaling proteins.

The rich sensory repertoire of *S. algae* is reflected in the diversity of the N-terminal sensory domains of c-di-GMP turnover proteins (Fig. 3 and 4). We have described here several previously uncharacterized ones, including three new MASE (membrane-associated sensor) domains (MASE6 to MASE8), a novel VUPS (vitamin uptake-like sensor) domain, a novel DGCcoil (DGC-associated coil) domain, and a new CSS_CxxC variant of the CSS-motif sensory domain (86) containing two additional Cys residues that presumably broaden its redox-sensing palette. Unraveling the full breadth of signals and protein-protein interactions of these novel domains will require extensive experimental investigation.

The analysis of the chemosensory pathways of *S. algae* revealed the existence of two conserved pathways: an F6 pathway involved in chemotaxis and an F7 pathway of unknown function. The resemblance of the *S. algae* and *P. aeruginosa* chemosensory systems is noteworthy. In contrast to *P. aeruginosa* strains, though, which exhibit a wide variability in terms of chemosensory content, *S. algae* strains have a rather narrow set of chemoreceptors. Inferences on signal recognition by *S. algae* chemoreceptors were made based on an array of *in silico* tools, including homology modeling. There are currently no experimental data on the chemoeffectors recognized by these chemoreceptors. This lack of knowledge represents a clear research need that extends to other marine bacteria, whose dissection is likely to yield valuable results from the ecophysiological perspective. Notably, genomic analysis of the acute enteritis isolates G1 and A41 revealed a putative integrative and conjugative element containing a putative aerotaxis receptor, a CheV homolog, two c-di-GMP turnover proteins, and a PAS domain-containing chemoreceptor similar to *P. aeruginosa* BdlA (biofilm dispersion locus A), known to be involved in virulence. As *S. algae* virulence and infectivity determinants in the human host are virtually uncharacterized, the analysis of this element may provide a valuable entry point for the study of *S. algae* infectivity and lifestyle regulation in the intestinal tract.

Altogether, we have used state-of-the-art methodologies to construct a comprehensive atlas of *S. algae* c-di-GMP and chemosensory systems, which establishes *S. algae* as a model organism for the analysis of signaling pathways that will serve as a reference for future genetic and functional studies in *Shewanella*.

MATERIALS AND METHODS

Bacterial strains and whole-genome sequencing. *Shewanella algae* strains (Table S1) were grown in LB medium at 37°C prior to DNA isolation with (i) the GenElute bacterial genomic DNA kit (Merck) for samples submitted for WGS with Illumina MiSeq or (ii) with the Genomic-tip 500/G kit (Qiagen) for samples submitted for Oxford Nanopore or PacBio sequencing, as detailed below.

(i) PacBio sequencing. Multiplexed sequencing libraries for strain 150735 were constructed using the SMRTbell Express template prep kit 2.0 and the Barcoded Overhang Adapter kit 8A (PacBio, Menlo Park, CA, USA). Libraries were sequenced on one Sequel SMRT cell according to the manufacturer's recommendations (PacBio).

(ii) Oxford Nanopore sequencing. Genomic DNA samples of strains A59, A291, and G1 were barcoded using the SQK-RBK004 Rapid Barcoding kit and sequenced on the MinION platform, flow cell version FLO-MIN106. Fast5 files were base called and demultiplexed using the Guppy software v4.2.2. Hybrid assemblies were performed using the Unicycler software v0.4.8.

(iii) Illumina sequencing. Sequencing libraries from all other 37 *S. algae* strains were prepared using the TruSeq Nano DNA library prep (Illumina, San Diego, CA, USA) and sequenced at the Center for Translational Microbiome Research, Karolinska Institute, on an Illumina MiSeq instrument with 2 × 300-bp paired-end reads. Sequencing reads were processed using the BACTpipe bacterial assembly and annotation pipeline v2.6.0 (https://github.com/ctmrbio/BACTpipe).

The long reads from the PacBio sequencing were assembled using HGAP4 included in SMRT Link v8 (PacBio). The contigs from the long-read assemblies were manually circularized and then linearized to start close to the origin of replication (at the start of the *dnaA* locus).

For each strain, contigs from the short-read assemblies were ordered according to their mapping to the type strain CECT 5071. Contigs spanning the *dnaA* locus were split 5′ of the *dnaA* gene. The sequence part of the contig containing *dnaA* was placed at the beginning of the assembly and the upstream part at the end of the matching contigs. Non-unique-mapping contigs and nonmapping contigs were added after the last matching contig.

Prokka (125) annotations of draft and complete genomes were used for primary bioinformatics; therefore, Prokka-generated locus tags were used for *in silico* analyses. Genome sequences were deposited in GenBank (BioProject PRJNA526057) under accession no. SAMN11083189 (CECT 5071) (39) and SAMN16670435 to SAMN16670475 (all other 41 strains), along with the other accession numbers provided in the "Data availability" section below, and annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (126). The correspondence between the Prokka gene locus tags, GenBank accession number genome locus tags, and GenBank accession numbers for relevant sequences is presented in Table S2 in the supplemental material.

Whole-genome sequence assembly comparisons and pangenome analysis. WGS assemblies were compared using Sourmash 4.2.1 (127). With a k-size of 31 and scaled setting of 1,000, genome sketches for each *S. algae* strain were subjected to an all-by-all comparison to evaluate their pairwise relative genomic distance (measured as the Jaccard index between two assemblies). Taxonomic assessment of each WGS MinHash sketch was achieved by querying each genome sketch against GTDB (128) r202 (available at https://osf.io/wxf9z/) to identify the best match based on Jaccard similarity via the search function of Sourmash. Code used to generate Fig. 1 is available at https://github.com/ctmrbio/Salgae_c-di-GMP_analysis. Pairwise comparisons between the WGS of the type strain CECT 5071 and the WGSs of all other 41 strains were also performed using the Genome-to-Genome Distance Calculator v2.1 (Leibniz Institute DSMZ) under the recommended settings (43), and pairwise digital DNA-DNA hybridization (dDDH) values were inferred. The threshold for species and subspecies delineation was set at DDH >70% and DDH >79%, respectively, as previously described (40, 43). Pangenome construction was performed using the Roary pipeline (129) with a BLASTp (130) identity threshold of 90%. The gene presence/absence content is available in Data Set S1, and GFF files including Prokka annotations are available at https://github.com/ctmrbio/Salgae_c-di-GMP_analysis.

Subsequent pangenome analyses using gene presence/absence data from Roary were carried out with the R package Pagoo 0.3.9 (131). EggNog mapper 2.1.6 (132) was used to generate KEGG orthology annotations for each homologous gene cluster identified by Roary via diamond queries against the eggNog v5.0 database (133). KEGG orthology annotations for each gene cluster were then mapped to their corresponding KEGG pathways by referencing the KEGG brite file br08901.keg and a KEGG pathway to KEGG orthology id mapping file (http://rest.kegg.jp/link/ko/pathway). Code used to generate Fig. 2 is available at https://github.com/ctmrbio/Salgae_c-di-GMP_analysis.

Other relevant phylogenetic analyses were performed with MEGAX (134). Genomic islands were predicted using IslandViewer 4 (135).

Bioinformatic analyses of cyclic di-GMP metabolizing proteins. Putative diguanylate cyclases (containing the GGDEF domain) and putative c-di-GMP phosphodiesterases (containing either EAL or HD-GYP domains) were identified in *S. algae* proteomes using searches with BLAST (130), ProSite (136), and HMMER (137). Verification of the occurrence of a GGDEF, EAL, and/or HD-GYP domain in candidate proteins was manually investigated by running the protein sequence through additional domain recognition programs such as CDD-Search (138), SMART (139), and Pfam (47). Poorly conserved domains were identified with HHpred (140). A search for orthologs of *S. algae* CECT 5071 c-di-GMP turnover proteins in other *S. algae* genomes was performed by BLASTp (130) followed by manual curation of the resulting hits.

For the initial multiple alignments, the entire protein sequences including truncated domains were aligned with ClustalX (141). The sequence alignment was then manually assessed and curated in GeneDoc (142) by removing gaps and/or adjacent domains. For structure-based alignments, the respective GGDEF, EAL, and HD-GYP domains were aligned against the reference proteins as indicated in the legends to the figures. Structure-based alignment was performed with ESPript 3.0 (143) using default parameters.

Cyclic di-GMP turnover protein maximum likelihood trees were produced using RaxML-NG (144) with default parameters (LG substitution matrix, gamma-distributed substitution rates, 100 bootstrap replicates) on the Vital-IT website at https://raxml-ng.vital-it.ch/. The trees including the domain predictions were visualized using iTol (145). GrapeTree (146) was used for the visualization of the phylogenetic relationships of GGDEF, EAL, and HD-GYP domains of *S. algae* c-di-GMP turnover proteins not encoded in the type strain chromosome.

The newly described sensor domains were inferred from uncharacterized N-terminal sequences preceding the GGDEF and/or EAL domains. The respective sequences from *S. algae* were used as queries in PSI-BLAST (130) and HMMer (137) searches, and representative sequences from diverse bacteria and archaea were selected for inclusion in the displayed alignments. The novelty of identified domains and/or their remote similarity to any known domains was checked by HHPred (140). Transmembrane segments and their membrane topology were predicted with TMHMM2.0 (147) and verified by checking the respective entries in UniProt and inspecting the distribution of charged residues according to the "positive-inside rule" (148). In addition, predicted membrane topology was visualized by Protter (149). The putative N-terminal coiled-coil domain in the WT_00119 protein was analyzed using such prediction programs as MultiCoil2 (75) and PCOILS and DeepCoil (76, 77), accessed through the MPI Bioinformatics Toolkit (140). The quaternary structure of this domain was predicted with LOGICOIL (78).

Identification and analysis of chemosensory systems. The identification of chemoreceptors followed a similar method to that described above for c-di-GMP turnover proteins. Thus, a chemoreceptor is defined as a protein that contains an MCP signaling domain identified by either SMART (139), ProSite (136), Pfam (47), or CDVist (150) followed by manual curation of the obtained hits. Transmembrane regions were identified using TMHMM2.0 (147). Chemoreceptor ligand-binding domains (LBDs) were classified according to Pfam (47). In the case in which a potential LBD (as defined by a domain flanked by two transmembrane regions) was not annotated in Pfam, homology modeling using Phyre2 (151)

was used to approximate the LBD. The chemosensory signaling proteins were extracted from the MIST 3.0 database (152).

Data availability. The genome sequences used in this study are deposited in GenBank (BioProject PRJNA526057) under the following accession numbers (for BioSample numbers, strain names are given in parentheses): complete genomes, SAMN16670464 (150735), SAMN16670468 (G1), SAMN16670438 (A59), SAMN16670445 (A291), SAMN11083189 (CECT 5071); draft genomes, SAMN16670460 (404), SAMN16670458 (789), SAMN16670449 (CCUG 48086), SAMN16670450 (CCUG 38646), SAMN16670457 (CCUG 56496), SAMN16670447 (CCUG 72638), SAMN16670456 (CCUG 72678), SAMN16670465 (97087), SAMN16670459 (CCUG 58400), SAMN16670452 (CCUG 20533), SAMN16670467 (5043), SAMN16670453 (CCUG 15259), SAMN16670455 (CCUG 526), SAMN16670437 (A58), SAMN16670436 (A57), SAMN16670475 (950570), SAMN16670462 (6F5), SAMN16670440 (A41), SAMN16670446 (A292), SAMN16670435 (A56), SAMN16670466 (159418), SAMN16670470 (590722), SAMN16670469 (669801), SAMN16670451 (CCUG 24987), SAMN16670448 (CCUG 50501), SAMN16670463 (28011), SAMN16670454 (CCUG 12945), SAMN16670442 (A93), SAMN16670441 (A65), SAMN16670443 (A94), SAMN16670439 (A60), SAMN16670444 (A97), SAMN16670471 (HUD-G3), SAMN16670461 (SF7), SAMN16670473 (HUD-H4), SAMN16670474 (HUD-I2), SAMN16670472 (HUD-D8).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, TIF file, 0.9 MB.
FIG S2, PDF file, 1.4 MB.
FIG S3, TIF file, 0.6 MB.
FIG S4, TIF file, 0.3 MB.
FIG S5, PDF file, 0.3 MB.
FIG S6, PDF file, 0.2 MB.
FIG S7, TIF file, 0.7 MB.
TABLE S1, DOCX file, 0.03 MB.
TABLE S2, DOCX file, 0.02 MB.
DATA SET S1, XLSX file, 2.3 MB.

ACKNOWLEDGMENTS

A.J.M.-R. is grateful to Volkan Özenci, José Gutiérrez-Fernández, and Lone Gram for the kind gift of *S. algae* strains. We are grateful to Igor Zhulin for the inclusion of the *S. algae* genomes in the MIST3 database.

Funding for this study was provided by grants from Stiftelsen Lars Hiertas Minne (grant FO2019-0293), Stiftelsen Längmanska Kulturfonden (grant BA20-0736), the Karolinska Institute Research Foundation (grant 2020-01556), Stiftelsen Anna och Gunnar Vidfelts fond för biologisk forskning (grant 2019-051-Vidfelts fond/SOJOH) and the Hans Dahlbergs Stiftelse för miljö och hälsa to A.J.M.-R. M.Y.G. was supported by the Intramural Research Program of the National Library of Medicine, U.S. National Institutes of Health. T.K. was supported by grants from the Junta de Andalucía (P18-FR-1621) and the Spanish Ministry of Economy and Competitiveness (PID2020-112612GB-I00). U.R. was supported by the Swedish Research Council for Natural Sciences and Engineering (2017-04465) and the Karolinska Institute.

REFERENCES

1. Hau HH, Gralnick JA. 2007. Ecology and biotechnology of the genus *Shewanella*. Annu Rev Microbiol 61:237–258. https://doi.org/10.1146/annurev.micro.61.080706.093257.
2. Lemaire ON, Mejean V, Iobbi-Nivol C. 2020. The *Shewanella* genus: ubiquitous organisms sustaining and preserving aquatic ecosystems. FEMS Microbiol Rev 44:155–170. https://doi.org/10.1093/femsre/fuz031.
3. Cha QQ, Ren XB, Sun YY, He XY, Su HN, Chen XL, Zhang YZ, Xie BB, Zhao LS, Song XY, Zhang XY. 2020. *Shewanella polaris* sp. nov., a psychrotolerant bacterium isolated from Arctic brown algae. Int J Syst Evol Microbiol 70:2096–2102. https://doi.org/10.1099/ijsem.0.004022.
4. Ivanova EP, Sawabe T, Gorshkova NM, Svetashev VI, Mikhailov VV, Nicolau DV, Christen R. 2001. *Shewanella japonica* sp. nov. Int J Syst Evol Microbiol 51:1027–1033. https://doi.org/10.1099/00207713-51-3-1027.
5. Naghoni A, Emtiazi G, Amoozegar MA, Cretoiu MS, Stal LJ, Etemadifar Z, Shahzadeh Fazeli SA, Bolhuis H. 2017. Microbial diversity in the hypersaline Lake Meyghan, Iran. Sci Rep 7:11522. https://doi.org/10.1038/s41598-017-11585-3.
6. Yang SH, Lee JH, Ryu JS, Kato C, Kim SJ. 2007. *Shewanella donghaensis* sp. nov., a psychrophilic, piezosensitive bacterium producing high levels of polyunsaturated fatty acid, isolated from deep-sea sediments. Int J Syst Evol Microbiol 57:208–212. https://doi.org/10.1099/ijs.0.64469-0.
7. Caro-Quintero A, Deng J, Auchtung J, Brettar I, Hofle MG, Klappenbach J, Konstantinidis KT. 2011. Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea. ISME J 5:131–140. https://doi.org/10.1038/ismej.2010.93.
8. Martín-Rodríguez AJ, Martin-Pujol O, Artiles-Campelo F, Bolanos-Rivero M, Römling U. 2017. *Shewanella* spp. infections in Gran Canaria, Spain: retrospective analysis of 31 cases and a literature review. JMM Case Rep 4:e005131.
9. Bauer MJ, Stone-Garza KK, Croom D, Andreoli C, Woodson P, Graf PCF, Maves RC. 2019. *Shewanella algae* infections in United States Naval

Special Warfare trainees. Open Forum Infect Dis 6:ofz442. https://doi.org/10.1093/ofid/ofz442.
10. Cai J, Chen H, Thompson KD, Li C. 2006. Isolation and identification of *Shewanella alga* and its pathogenic effects on post-larvae of abalone *Haliotis diversicolor supertexta*. J Fish Dis 29:505–508. https://doi.org/10.1111/j.1365-2761.2006.00732.x.
11. Chen C, Hu C, Chen X, Zhang L. 2003. Identification and characterization of *Shewanella algae* as a novel pathogen of ulcer disease of fish *Scinenops ocellata*. Oceanol Limnol Sinica 34:1–8.
12. Erfanmanesh A, Beikzadeh B, Aziz Mohseni F, Nikaein D, Mohajerfar T. 2019. Ulcerative dermatitis in barramundi due to coinfection with *Streptococcus iniae* and *Shewanella algae*. Dis Aquat Organ 134:89–97. https://doi.org/10.3354/dao03363.
13. Konstantinidis KT, Serres MH, Romine MF, Rodrigues JL, Auchtung J, McCue LA, Lipton MS, Obraztsova A, Giometti CS, Nealson KH, Fredrickson JK, Tiedje JM. 2009. Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. Proc Natl Acad Sci U S A 106:15909–15914. https://doi.org/10.1073/pnas.0902000106.
14. Zhong C, Han M, Yu S, Yang P, Li H, Ning K. 2018. Pan-genome analyses of 24 *Shewanella* strains re-emphasize the diversification of their functions yet evolutionary dynamics of metal-reducing pathway. Biotechnol Biofuels 11:193. https://doi.org/10.1186/s13068-018-1201-1.
15. Martín-Rodríguez AJ, Reyes-Darias JA, Martin-Mora D, Gonzalez JM, Krell T, Römling U. 2021. Reduction of alternative electron acceptors drives biofilm formation in *Shewanella algae*. NPJ Biofilms Microbiomes 7:9. https://doi.org/10.1038/s41522-020-00177-1.
16. Römling U, Galperin MY, Gomelsky M. 2013. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol Mol Biol Rev 77:1–52. https://doi.org/10.1128/MMBR.00043-12.
17. Martín-Rodríguez AJ, Römling U. 2017. Nucleotide second messenger signaling as a target for the control of bacterial biofilm formation. Curr Top Med Chem 2017:1928–1944. https://doi.org/10.2174/1568026617666170105144424.
18. Hengge R, Galperin MY, Ghigo JM, Gomelsky M, Green J, Hughes KT, Jenal U, Landini P. 2016. Systematic nomenclature for GGDEF and EAL domain-containing cyclic di-GMP turnover proteins of *Escherichia coli*. J Bacteriol 198:7–11. https://doi.org/10.1128/JB.00424-15.
19. Sarenko O, Klauck G, Wilke FM, Pfiffer V, Richter AM, Herbst S, Kaever V, Hengge R. 2017. More than enzymes that make or break cyclic di-GMP—local signaling in the interactome of GGDEF/EAL domain proteins of *Escherichia coli*. mBio 8:e01639-17. https://doi.org/10.1128/mBio.01639-17.
20. Richter AM, Possling A, Malysheva N, Yousef KP, Herbst S, von Kleist M, Hengge R. 2020. Local c-di-GMP signaling in the control of synthesis of the *E. coli* biofilm exopolysaccharide pEtN-cellulose. J Mol Biol 432:4576–4595. https://doi.org/10.1016/j.jmb.2020.06.006.
21. Kader A, Simm R, Gerstel U, Morr M, Römling U. 2006. Hierarchical involvement of various GGDEF domain proteins in rdar morphotype development of *Salmonella enterica* serovar Typhimurium. Mol Microbiol 60:602–616. https://doi.org/10.1111/j.1365-2958.2006.05123.x.
22. Galperin MY. 2005. A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. BMC Microbiol 5:35. https://doi.org/10.1186/1471-2180-5-35.
23. Galperin MY, Higdon R, Kolker E. 2010. Interplay of heritage and habitat in the distribution of bacterial signal transduction systems. Mol Biosyst 6:721–728. https://doi.org/10.1039/b908047c.
24. Chao L, Rakshe S, Leff M, Spormann AM. 2013. PdeB, a cyclic di-GMP-specific phosphodiesterase that regulates *Shewanella oneidensis* MR-1 motility and biofilm formation. J Bacteriol 195:3827–3833. https://doi.org/10.1128/JB.00498-13.
25. Nisbett LM, Binnenkade L, Bacon B, Hossain S, Kotloski NJ, Brutinel ED, Hartmann R, Drescher K, Arora DP, Muralidharan S, Thormann KM, Gralnick JA, Boon EM. 2019. NosP signaling modulates the NO/H-NOX-mediated multicomponent c-di-GMP network and biofilm formation in *Shewanella oneidensis*. Biochemistry 58:4827–4841. https://doi.org/10.1021/acs.biochem.9b00706.
26. Pecina A, Schwan M, Blagotinsek V, Rick T, Kluber P, Leonhard T, Bange G, Thormann KM. 2021. The stand-alone PilZ-domain protein MotL specifically regulates the activity of the secondary lateral flagellar system in *Shewanella putrefaciens*. Front Microbiol 12:668892. https://doi.org/10.3389/fmicb.2021.668892.
27. Yildiz FH, Visick KL. 2009. *Vibrio* biofilms: so much the same yet so different. Trends Microbiol 17:109–118. https://doi.org/10.1016/j.tim.2008.12.004.
28. Conner JG, Zamorano-Sanchez D, Park JH, Sondermann H, Yildiz FH. 2017. The ins and outs of cyclic di-GMP signaling in *Vibrio cholerae*. Curr Opin Microbiol 36:20–29. https://doi.org/10.1016/j.mib.2017.01.002.
29. Wuichet K, Zhulin IB. 2010. Origins and diversification of a complex signal transduction system in prokaryotes. Sci Signal 3:ra50. https://doi.org/10.1126/scisignal.2000724.
30. Porter SL, Wadhams GH, Armitage JP. 2011. Signal processing in complex chemotaxis pathways. Nat Rev Microbiol 9:153–165. https://doi.org/10.1038/nrmicro2505.
31. Baker MD, Wolanin PM, Stock JB. 2006. Signal transduction in bacterial chemotaxis. Bioessays 28:9–22. https://doi.org/10.1002/bies.20343.
32. Salah Ud-Din AIM, Roujeinikova A. 2017. Methyl-accepting chemotaxis proteins: a core sensing element in prokaryotes and archaea. Cell Mol Life Sci 74:3293–3303. https://doi.org/10.1007/s00018-017-2514-0.
33. Krell T, Lacal J, Munoz-Martinez F, Reyes-Darias JA, Cadirci BH, Garcia-Fontana C, Ramos JL. 2011. Diversity at its best: bacterial taxis. Environ Microbiol 13:1115–1124. https://doi.org/10.1111/j.1462-2920.2010.02383.x.
34. Ortega A, Zhulin IB, Krell T. 2017. Sensory repertoire of bacterial chemoreceptors. Microbiol Mol Biol Rev 81:e00033-17. https://doi.org/10.1128/MMBR.00033-17.
35. Boyeldieu A, Ali Chaouche A, Mejean V, Jourlin-Castelli C. 2021. Combining two optimized and affordable methods to assign chemoreceptors to a specific signal. Anal Biochem 620:114139. https://doi.org/10.1016/j.ab.2021.114139.
36. Cheng L, Min D, Liu DF, Li WW, Yu HQ. 2019. Sensing and approaching toxic arsenate by *Shewanella putrefaciens* CN-32. Environ Sci Technol 53:14604–14611. https://doi.org/10.1021/acs.est.9b05890.
37. Starwalt-Lee R, El-Naggar MY, Bond DR, Gralnick JA. 2021. Electrolocation? The evidence for redox-mediated taxis in *Shewanella oneidensis*. Mol Microbiol 115:1069–1079. https://doi.org/10.1111/mmi.14647.
38. Ortega DR, Yang W, Subramanian P, Mann P, Kjær A, Chen S, Watts KJ, Pirbadian S, Collins DA, Kooger R, Kalyuzhnaya MG, Ringgaard S, Briegel A, Jensen GJ. 2020. Repurposing a chemosensory macromolecular machine. Nat Commun 11:2041. https://doi.org/10.1038/s41467-020-15736-5.
39. Tellgren-Roth C, Thorell K, Galperin MY, Krell T, Römling U, Sjöling A, Martín-Rodríguez AJ. 2021. Complete genome sequence and methylome of the type strain of *Shewanella algae*. Microbiol Resour Announc 10:e00559-21. https://doi.org/10.1128/MRA.00559-21.
40. Thorell K, Meier-Kolthoff JP, Sjoling A, Martin-Rodriguez AJ. 2019. Whole-genome sequencing redefines *Shewanella* taxonomy. Front Microbiol 10:1861. https://doi.org/10.3389/fmicb.2019.01861.
41. Martín-Rodríguez AJ, Suárez-Mesa A, Artiles-Campelo F, Römling U, Hernandez M. 2019. Multilocus sequence typing of *Shewanella algae* isolates identifies disease-causing *Shewanella chilikensis* strain 6I4. FEMS Microbiol Ecol 95:fiy210. https://doi.org/10.1093/femsec/fiy210.
42. Szeinbaum N, Kellum CE, Glass JB, Janda JM, DiChristina TJ. 2018. Whole-genome sequencing reveals that *Shewanella haliotis* Kim et al. 2007 can be considered a later heterotypic synonym of *Shewanella algae* Simidu et al. 1990. Int J Syst Evol Microbiol 68:1356–1360. https://doi.org/10.1099/ijsem.0.002678.
43. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60. https://doi.org/10.1186/1471-2105-14-60.
44. Martín-Rodríguez AJ, Villion K, Yilmaz-Turan S, Vilaplana F, Sjöling Å, Römling U. 2021. Regulation of colony morphology and biofilm formation in *Shewanella algae*. Microb Biotechnol 14:1183–1200. https://doi.org/10.1111/1751-7915.13788.
45. Thormann KM, Duttler S, Saville RM, Hyodo M, Shukla S, Hayakawa Y, Spormann AM. 2006. Control of formation and cellular detachment from *Shewanella oneidensis* MR-1 biofilms by cyclic di-GMP. J Bacteriol 188:2681–2691. https://doi.org/10.1128/JB.188.7.2681-2691.2006.
46. Wu C, Cheng YY, Yin H, Song XN, Li WW, Zhou XX, Zhao LP, Tian LJ, Han JC, Yu HQ. 2013. Oxygen promotes biofilm formation of *Shewanella putrefaciens* CN32 through a diguanylate cyclase and an adhesin. Sci Rep 3:1945. https://doi.org/10.1038/srep01945.
47. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: the protein families database in 2021. Nucleic Acids Res 49:D412–D419. https://doi.org/10.1093/nar/gkaa913.
48. Christen B, Christen M, Paul R, Schmid F, Folcher M, Jenoe P, Meuwly M, Jenal U. 2006. Allosteric control of cyclic di-GMP signaling. J Biol Chem 281:32015–32024. https://doi.org/10.1074/jbc.M603589200.

49. Galperin MY. 2006. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. J Bacteriol 188:4169–4182. https://doi.org/10.1128/JB.01887-05.
50. Jonas K, Tomenius H, Romling U, Georgellis D, Melefors O. 2006. Identification of YhdA as a regulator of the *Escherichia coli* carbon storage regulation system. FEMS Microbiol Lett 264:232–237. https://doi.org/10.1111/j.1574-6968.2006.00457.x.
51. Suzuki K, Babitzke P, Kushner SR, Romeo T. 2006. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. Genes Dev 20:2605–2617. https://doi.org/10.1101/gad.1461606.
52. Newell PD, Monds RD, O'Toole GA. 2009. LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by *Pseudomonas fluorescens* Pf0-1. Proc Natl Acad Sci U S A 106:3461–3466. https://doi.org/10.1073/pnas.0808933106.
53. Collins AJ, Smith TJ, Sondermann H, O'Toole GA. 2020. From input to output: the Lap/c-di-GMP biofilm regulatory circuit. Annu Rev Microbiol 74:607–631. https://doi.org/10.1146/annurev-micro-011520-094214.
54. Fineran PC, Williamson NR, Lilley KS, Salmond GP. 2007. Virulence and prodigiosin antibiotic biosynthesis in *Serratia* are regulated pleiotropically by the GGDEF/EAL domain protein, PigX. J Bacteriol 189:7653–7662. https://doi.org/10.1128/JB.00671-07.
55. El Mouali Y, Kim H, Ahmad I, Brauner A, Liu Y, Skurnik M, Galperin MY, Römling U. 2017. Stand-alone EAL domain proteins form a distinct subclass of EAL proteins involved in regulation of cell motility and biofilm formation in enterobacteria. J Bacteriol 199:e00179-17. https://doi.org/10.1128/JB.00179-17.
56. Römling U. 2020. Cyclic di-GMP signaling in *Salmonella enterica* serovar Typhimurium, p 395–425. *In* Chou SH, Guiliani N, Lee VT, Römling U (ed), Microbial cyclic di-nucleotide signaling. Springer, Cham, Switzerland.
57. Galperin MY, Natale DA, Aravind L, Koonin EV. 1999. A specialized version of the HD hydrolase domain implicated in signal transduction. J Mol Microbiol Biotechnol 1:303–305.
58. Bellini D, Caly DL, McCarthy Y, Bumann M, An SQ, Dow JM, Ryan RP, Walsh MA. 2014. Crystal structure of an HD-GYP domain cyclic-di-GMP phosphodiesterase reveals an enzyme with a novel trinuclear catalytic iron centre. Mol Microbiol 91:26–38. https://doi.org/10.1111/mmi.12447.
59. Gao J, Tao J, Liang W, Zhao M, Du X, Cui S, Duan H, Kan B, Su X, Jiang Z. 2015. Identification and characterization of phosphodiesterases that specifically degrade 3'3'-cyclic GMP-AMP. Cell Res 25:539–550. https://doi.org/10.1038/cr.2015.40.
60. Deng MJ, Tao J, E C, Ye ZY, Jiang Z, Yu J, Su XD. 2018. Novel mechanism for cyclic dinucleotide degradation revealed by structural studies of *Vibrio* phosphodiesterase V-cGAP3. J Mol Biol 430:5080–5093. https://doi.org/10.1016/j.jmb.2018.10.010.
61. Galperin MY, Chou SH. 2022. Sequence conservation, domain architectures, and phylogenetic distribution of the HD-GYP type c-di-GMP phosphodiesterases. J Bacteriol 204:e00561-21. https://doi.org/10.1128/jb.00561-21.
62. Chen HJ, Li N, Luo Y, Jiang YL, Zhou CZ, Chen Y, Li Q. 2018. The GDP-switched GAF domain of DcpA modulates the concerted synthesis/hydrolysis of c-di-GMP in *Mycobacterium smegmatis*. Biochem J 475:1295–1308. https://doi.org/10.1042/BCJ20180079.
63. Ereno-Orbea J, Oyenarte I, Martinez-Cruz LA. 2013. CBS domains: ligand binding sites and conformational variability. Arch Biochem Biophys 540:70–81. https://doi.org/10.1016/j.abb.2013.10.008.
64. Upadhyay AA, Fleetwood AD, Adebali O, Finn RD, Zhulin IB. 2016. Cache domains that are homologous to, but different from PAS domains comprise the largest superfamily of extracellular sensors in prokaryotes. PLoS Comput Biol 12:e1004862. https://doi.org/10.1371/journal.pcbi.1004862.
65. Galperin MY, Gaidenko TA, Mulkidjanian AY, Nakano M, Price CW. 2001. MHYT, a new integral membrane sensor domain. FEMS Microbiol Lett 205:17–23. https://doi.org/10.1111/j.1574-6968.2001.tb10919.x.
66. Li Y, Heine S, Entian M, Sauer K, Frankenberg-Dinkel N. 2013. NO-induced biofilm dispersion in *Pseudomonas aeruginosa* is mediated by an MHYT domain-coupled phosphodiesterase. J Bacteriol 195:3531–3542. https://doi.org/10.1128/JB.01156-12.
67. Anantharaman V, Aravind L. 2003. Application of comparative genomics in the identification and analysis of novel families of membrane-associated receptors in bacteria. BMC Genomics 4:34. https://doi.org/10.1186/1471-2164-4-34.
68. Nikolskaya AN, Mulkidjanian AY, Beech IB, Galperin MY. 2003. MASE1 and MASE2: two novel integral membrane sensory domains. J Mol Microbiol Biotechnol 5:11–16. https://doi.org/10.1159/000068720.
69. Galperin MY, Nikolskaya AN. 2007. Identification of sensory and signal-transducing domains in two-component signaling systems. Methods Enzymol 422:47–74. https://doi.org/10.1016/S0076-6879(06)22003-2.
70. Lacey M, Agasing A, Lowry R, Green J. 2013. Identification of the YfgF MASE1 domain as a modulator of bacterial responses to aspartate. Open Biol 3:130046. https://doi.org/10.1098/rsob.130046.
71. Pfiffer V, Sarenko O, Possling A, Hengge R. 2019. Genetic dissection of *Escherichia coli*'s master diguanylate cyclase DgcE: role of the N-terminal MASE1 domain and direct signal input from a GTPase partner system. PLoS Genet 15:e1008059. https://doi.org/10.1371/journal.pgen.1008059.
72. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Res 49:D274–D281. https://doi.org/10.1093/nar/gkaa1018.
73. Zallot R, Yuan Y, de Crecy-Lagard V. 2017. The *Escherichia coli* COG1738 member YhhQ is involved in 7-cyanodeazaguanine ($preQ_0$) transport. Biomolecules 7:12. https://doi.org/10.3390/biom7010012.
74. Yuan Y, Zallot R, Grove TL, Payan DJ, Martin-Verstraete I, Sepic S, Balamkundu S, Neelakandan R, Gadi VK, Liu CF, Swairjo MA, Dedon PC, Almo SC, Gerlt JA, de Crecy-Lagard V. 2019. Discovery of novel bacterial queuine salvage enzymes and pathways in human pathogens. Proc Natl Acad Sci U S A 116:19126–19135. https://doi.org/10.1073/pnas.1909604116.
75. Trigg J, Gutwin K, Keating AE, Berger B. 2011. Multicoil2: predicting coiled coils and their oligomerization states from sequence in the twilight zone. PLoS One 6:e23519. https://doi.org/10.1371/journal.pone.0023519.
76. Gruber M, Söding J, Lupas AN. 2006. Comparative analysis of coiled-coil prediction methods. J Struct Biol 155:140–145. https://doi.org/10.1016/j.jsb.2006.03.009.
77. Ludwiczak J, Winski A, Szczepaniak K, Alva V, Dunin-Horkawicz S. 2019. DeepCoil—a fast and accurate prediction of coiled-coil domains in protein sequences. Bioinformatics 35:2790–2795. https://doi.org/10.1093/bioinformatics/bty1062.
78. Vincent TL, Green PJ, Woolfson DN. 2013. LOGICOIL—multi-state prediction of coiled-coil oligomeric state. Bioinformatics 29:69–76. https://doi.org/10.1093/bioinformatics/bts648.
79. Ulrich LE, Zhulin IB. 2005. Four-helix bundle: a ubiquitous sensory module in prokaryotic signal transduction. Bioinformatics 21:iii45–iii48. https://doi.org/10.1093/bioinformatics/bti1204.
80. Beyhan S, Odell LS, Yildiz FH. 2008. Identification and characterization of cyclic diguanylate signaling systems controlling rugosity in *Vibrio cholerae*. J Bacteriol 190:7392–7405. https://doi.org/10.1128/JB.00564-08.
81. Wei Q, Leclercq S, Bhasme P, Xu A, Zhu B, Zhang Y, Zhang M, Wang S, Ma LZ. 2019. Diguanylate cyclases and phosphodiesterases required for basal-level c-di-GMP in *Pseudomonas aeruginosa* as revealed by systematic phylogenetic and transcriptomic analyses. Appl Environ Microbiol 85:e01194-19. https://doi.org/10.1128/AEM.01194-19.
82. Pineda-Molina E, Reyes-Darias JA, Lacal J, Ramos JL, Garcia-Ruiz JM, Gavira JA, Krell T. 2012. Evidence for chemoreceptors with bimodular ligand-binding regions harboring two signal-binding sites. Proc Natl Acad Sci U S A 109:18926–18931. https://doi.org/10.1073/pnas.1201400109.
83. Ortega A, Krell T. 2014. The HBM domain: introducing bimodularity to bacterial sensing. Protein Sci 23:332–336. https://doi.org/10.1002/pro.2410.
84. Elgamoudi BA, Andrianova EP, Shewell LK, Day CJ, King RM, Taha Rahman H, Hartley-Tassell LE, Zhulin IB, Korolik V. 2021. The *Campylobacter jejuni* chemoreceptor Tlp10 has a bimodal ligand-binding domain and specificity for multiple classes of chemoeffectors. Sci Signal 14:eabc8521. https://doi.org/10.1126/scisignal.abc8521.
85. Moore JO, Hendrickson WA. 2009. Structural analysis of sensor domains from the TMAO-responsive histidine kinase receptor TorS. Structure 17:1195–1204. https://doi.org/10.1016/j.str.2009.07.015.
86. Herbst S, Lorkowski M, Sarenko O, Nguyen TKL, Jaenicke T, Hengge R. 2018. Transmembrane redox control and proteolysis of PdeC, a novel type of c-di-GMP phosphodiesterase. EMBO J 37:e97825. https://doi.org/10.15252/embj.201797825.
87. Hoffman LR, D'Argenio DA, MacCoss MJ, Zhang Z, Jones RA, Miller SI. 2005. Aminoglycoside antibiotics induce bacterial biofilm formation. Nature 436:1171–1175. https://doi.org/10.1038/nature03912.
88. Fomenko DE, Gladyshev VN. 2003. Identity and functions of CxxC-derived motifs. Biochemistry 42:11214–11225. https://doi.org/10.1021/bi034459s.

89. Denoncin K, Collet J-F. 2013. Disulfide bond formation in the bacterial periplasm: major achievements and challenges ahead. Antioxid Redox Signal 19:63–71. https://doi.org/10.1089/ars.2012.4864.
90. Corral-Lugo A, De la Torre J, Matilla MA, Fernandez M, Morel B, Espinosa-Urgel M, Krell T. 2016. Assessment of the contribution of chemoreceptor-based signalling to biofilm formation. Environ Microbiol 18:3355–3372. https://doi.org/10.1111/1462-2920.13170.
91. Huang Z, Wang YH, Zhu HZ, Andrianova EP, Jiang CY, Li D, Ma L, Feng J, Liu ZP, Xiang H, Zhulin IB, Liu SJ. 2019. Cross talk between chemosensory pathways that modulate chemotaxis and biofilm formation. mBio 10: e02876-18. https://doi.org/10.1128/mBio.02876-18.
92. Morgan R, Kohn S, Hwang SH, Hassett DJ, Sauer K. 2006. BdlA, a chemotaxis regulator essential for biofilm dispersion in *Pseudomonas aeruginosa*. J Bacteriol 188:7335–7343. https://doi.org/10.1128/JB.00599-06.
93. Cerna-Vargas JP, Santamaria-Hernando S, Matilla MA, Rodriguez-Herva JJ, Daddaoua A, Rodriguez-Palenzuela P, Krell T, Lopez-Solanilla E. 2019. Chemoperception of specific amino acids controls phytopathogenicity in *Pseudomonas syringae* pv. *tomato*. mBio 10:e01868-19. https://doi.org/10.1128/mBio.01868-19.
94. Xu A, Wang D, Wang Y, Zhang L, Xie Z, Cui Y, Bhamse P, Yu H, Zhang XX, Li D, Ma LZ. 2021. Mutations in surface-sensing receptor WspA lock the Wsp signal transduction system into a constitutively active state. Environ Microbiol https://doi.org/10.1111/1462-2920.15763.
95. Hickman JW, Tifrea DF, Harwood CS. 2005. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. Proc Natl Acad Sci U S A 102:14422–14427. https://doi.org/10.1073/pnas.0507170102.
96. Moscoso JA, Mikkelsen H, Heeb S, Williams P, Filloux A. 2011. The *Pseudomonas aeruginosa* sensor RetS switches type III and type VI secretion via c-di-GMP signalling. Environ Microbiol 13:3128–3138. https://doi.org/10.1111/j.1462-2920.2011.02595.x.
97. Boyeldieu A, Ali Chaouche A, Ba M, Honore FA, Mejean V, Jourlin-Castelli C. 2020. The phosphorylated regulator of chemotaxis is crucial throughout biofilm biogenesis in *Shewanella oneidensis*. NPJ Biofilms Microbiomes 6:54. https://doi.org/10.1038/s41522-020-00165-5.
98. Suchanek VM, Esteban-Lopez M, Colin R, Besharova O, Fritz K, Sourjik V. 2020. Chemotaxis and cyclic-di-GMP signalling control surface attachment of Escherichia coli. Mol Microbiol 113:728–739. https://doi.org/10.1111/mmi.14438.
99. Matilla MA, Martin-Mora D, Gavira JA, Krell T. 2021. *Pseudomonas aeruginosa* as a model to study chemosensory pathway signaling. Microbiol Mol Biol Rev 85:e00151-20. https://doi.org/10.1128/MMBR.00151-20.
100. Garcia-Fontana C, Vilchez JI, Gonzalez-Requena M, Gonzalez-Lopez J, Krell T, Matilla MA, Manzanera M. 2019. The involvement of McpB chemoreceptor from *Pseudomonas aeruginosa* PAO1 in virulence. Sci Rep 9: 13166. https://doi.org/10.1038/s41598-019-49697-7.
101. Güvener ZT, Tifrea DF, Harwood CS. 2006. Two different *Pseudomonas aeruginosa* chemosensory signal transduction complexes localize to cell poles and form and remould in stationary phase. Mol Microbiol 61: 106–118. https://doi.org/10.1111/j.1365-2958.2006.05218.x.
102. Gumerov VM, Andrianova EP, Zhulin IB. 2021. Diversity of bacterial chemosensory systems. Curr Opin Microbiol 61:42–50. https://doi.org/10.1016/j.mib.2021.01.016.
103. Lacal J, Garcia-Fontana C, Munoz-Martinez F, Ramos JL, Krell T. 2010. Sensing of environmental signals: classification of chemoreceptors according to the size of their ligand binding regions. Environ Microbiol 12:2873–2884. https://doi.org/10.1111/j.1462-2920.2010.02325.x.
104. Matilla MA, Krell T. 2017. Chemoreceptor-based signal sensing. Curr Opin Biotechnol 45:8–14. https://doi.org/10.1016/j.copbio.2016.11.021.
105. Matilla MA, Martin-Mora D, Krell T. 2020. The use of isothermal titration calorimetry to unravel chemotactic signalling mechanisms. Environ Microbiol 22:3005–3019. https://doi.org/10.1111/1462-2920.15035.
106. Hida A, Oku S, Nakashimada Y, Tajima T, Kato J. 2017. Identification of boric acid as a novel chemoattractant and elucidation of its chemoreceptor in *Ralstonia pseudosolanacearum* Ps29. Sci Rep 7:8609. https://doi.org/10.1038/s41598-017-09176-3.
107. Ni B, Huang Z, Fan Z, Jiang CY, Liu SJ. 2013. *Comamonas testosteroni* uses a chemoreceptor for tricarboxylic acid cycle intermediates to trigger chemotactic responses towards aromatic compounds. Mol Microbiol 90:813–823. https://doi.org/10.1111/mmi.12400.
108. Li Y, Liang J, Yang S, Yao J, Chen K, Yang L, Zheng W, Tian Y. 2021. Finding novel chemoreceptors that specifically sense and trigger chemotaxis toward polycyclic aromatic hydrocarbons in *Novosphingobium pentaromativorans* US6-1. J Hazard Mater 416:126246. https://doi.org/10.1016/j.jhazmat.2021.126246.
109. Shu CJ, Ulrich LE, Zhulin IB. 2003. The NIT domain: a predicted nitrate-responsive module in bacterial sensory receptors. Trends Biochem Sci 28: 121–124. https://doi.org/10.1016/S0968-0004(03)00032-X.
110. Martin-Mora D, Ortega A, Matilla MA, Martinez-Rodriguez S, Gavira JA, Krell T. 2019. The molecular mechanism of nitrate chemotaxis via direct ligand binding to the PilJ domain of McpN. mBio 10:e02334-18. https://doi.org/10.1128/mBio.02334-18.
111. Parkinson JS, Hazelbauer GL, Falke JJ. 2015. Signaling and sensory adaptation in *Escherichia coli* chemoreceptors: 2015 update. Trends Microbiol 23:257–266. https://doi.org/10.1016/j.tim.2015.03.003.
112. Muok AR, Ortega DR, Kurniyati K, Yang W, Maschmann ZA, Sidi Mabrouk A, Li C, Crane BR, Briegel A. 2020. Atypical chemoreceptor arrays accommodate high membrane curvature. Nat Commun 11:5763. https://doi.org/10.1038/s41467-020-19628-6.
113. Strube K, de Vries S, Cramm R. 2007. Formation of a dinitrosyl iron complex by NorA, a nitric oxide-binding di-iron protein from *Ralstonia eutropha* H16. J Biol Chem 282:20292–20300. https://doi.org/10.1074/jbc.M702003200.
114. Xiong J, Kurtz DM, Jr, Ai J, Sanders-Loehr J. 2000. A hemerythrin-like domain in a bacterial chemotaxis protein. Biochemistry 39:5117–5125. https://doi.org/10.1021/bi992796o.
115. Draper J, Karplus K, Ottemann KM. 2011. Identification of a chemoreceptor zinc-binding domain common to cytoplasmic bacterial chemoreceptors. J Bacteriol 193:4338–4345. https://doi.org/10.1128/JB.05140-11.
116. Collins KD, Andermann TM, Draper J, Sanders L, Williams SM, Araghi C, Ottemann KM. 2016. The *Helicobacter pylori* CZB cytoplasmic chemoreceptor TlpD forms an autonomous polar chemotaxis signaling complex that mediates a tactic response to oxidative stress. J Bacteriol 198:1563–1575. https://doi.org/10.1128/JB.00071-16.
117. Perkins A, Tudorica DA, Amieva MR, Remington SJ, Guillemin K. 2019. *Helicobacter pylori* senses bleach (HOCl) as a chemoattractant using a cytosolic chemoreceptor. PLoS Biol 17:e3000395. https://doi.org/10.1371/journal.pbio.3000395.
118. Hong Y, Huang Z, Guo L, Ni B, Jiang CY, Li XJ, Hou YJ, Yang WS, Wang DC, Zhulin IB, Liu SJ, Li DF. 2019. The ligand-binding domain of a chemoreceptor from *Comamonas testosteroni* has a previously unknown homotrimeric structure. Mol Microbiol 112:906–917. https://doi.org/10.1111/mmi.14326.
119. Alexander RP, Zhulin IB. 2007. Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. Proc Natl Acad Sci U S A 104:2885–2890. https://doi.org/10.1073/pnas.0609359104.
120. Ortega DR, Fleetwood AD, Krell T, Harwood CS, Jensen GJ, Zhulin IB. 2017. Assigning chemoreceptors to chemosensory pathways in *Pseudomonas aeruginosa*. Proc Natl Acad Sci U S A 114:12809–12814. https://doi.org/10.1073/pnas.1708842114.
121. Garcia D, Orillard E, Johnson MS, Watts KJ. 2017. Gas sensing and signaling in the PAS-heme domain of the *Pseudomonas aeruginosa* Aer2 receptor. J Bacteriol 199:e00003-17. https://doi.org/10.1128/JB.00003-17.
122. Greer-Phillips SE, Sukomon N, Chua TK, Johnson MS, Crane BR, Watts KJ. 2018. The Aer2 receptor from *Vibrio cholerae* is a dual PAS-heme oxygen sensor. Mol Microbiol 109:209–224. https://doi.org/10.1111/mmi.13978.
123. Garvis S, Munder A, Ball G, de Bentzmann S, Wiehlmann L, Ewbank JJ, Tummler B, Filloux A. 2009. *Caenorhabditis elegans* semi-automated liquid screen reveals a specialized role for the chemotaxis gene *cheB2* in *Pseudomonas aeruginosa* virulence. PLoS Pathog 5:e1000540. https://doi.org/10.1371/journal.ppat.1000540.
124. Matilla MA, Krell T. 2018. The effect of bacterial chemotaxis on host infection and pathogenicity. FEMS Microbiol Rev 42:fux052. https://doi.org/10.1093/femsre/fux052.
125. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
126. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
127. Brown CT, Irber L. 2016. sourmash: a library for MinHash sketching of DNA. JOSS 1:27. https://doi.org/10.21105/joss.00027.
128. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res 50:D785–D794. https://doi.org/10.1093/nar/gkab776.

129. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics/btv421.
130. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST—a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.
131. Ferrés I, Iraola G. 2021. An object-oriented framework for evolutionary pangenome analysis. Cell Rep Methods 1:100085. https://doi.org/10.1016/j.crmeth.2021.100085.
132. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 38:5825–5829. https://doi.org/10.1093/molbev/msab293.
133. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309–D314. https://doi.org/10.1093/nar/gky1085.
134. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096.
135. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, Brinkman FSL, Simon Fraser University Research Computing Group. 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.
136. Sigrist CJ, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. 2013. New and continuing developments at PROSITE. Nucleic Acids Res 41:D344–D347. https://doi.org/10.1093/nar/gks1067.
137. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR. 2015. HMMER web server: 2015 update. Nucleic Acids Res 43:W30–W38. https://doi.org/10.1093/nar/gkv397.
138. Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32:W327–W331. https://doi.org/10.1093/nar/gkh454.
139. Letunic I, Khedkar S, Bork P. 2021. SMART: recent updates, new developments and status in 2020. Nucleic Acids Res 49:D458–D460. https://doi.org/10.1093/nar/gkaa937.
140. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. J Mol Biol 430:2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007.
141. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882. https://doi.org/10.1093/nar/25.24.4876.
142. Nicholas KB, Nicholas HB, Jr, Deerfield DW. 1997. GeneDoc: analysis and visualization of genetic variation. EMB News 14:14.
143. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320–W324. https://doi.org/10.1093/nar/gku316.
144. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455. https://doi.org/10.1093/bioinformatics/btz305.
145. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239.
146. Zhou Z, Alikhan NF, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, Carrico JA, Achtman M. 2018. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res 28:1395–1404. https://doi.org/10.1101/gr.232397.117.
147. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.
148. von Heijne G. 1989. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. Nature 341:456–458. https://doi.org/10.1038/341456a0.
149. Omasits U, Ahrens CH, Muller S, Wollscheid B. 2014. Protter: interactive protein feature visualization and integration with experimental proteomic data. Bioinformatics 30:884–886. https://doi.org/10.1093/bioinformatics/btt607.
150. Adebali O, Ortega DR, Zhulin IB. 2015. CDvist: a webserver for identification and visualization of conserved domains in protein sequences. Bioinformatics 31:1475–1477. https://doi.org/10.1093/bioinformatics/btu836.
151. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053.
152. Gumerov VM, Ortega DR, Adebali O, Ulrich LE, Zhulin IB. 2020. MiST 3.0: an updated microbial signal transduction database with an emphasis on chemosensory systems. Nucleic Acids Res 48:D459–D464. https://doi.org/10.1093/nar/gkz988.