

RESEARCH

Open Access



# Patterns of abundance, chromosomal localization, and domain organization among c-di-GMP-metabolizing genes revealed by comparative genomics of five alphaproteobacterial orders

Sonja Koppenhöfer and Andrew S. Lang\*

## Abstract

**Background:** Bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) is a bacterial second messenger that affects diverse processes in different bacteria, including the cell cycle, motility, and biofilm formation. Its cellular levels are controlled by the opposing activities of two types of enzymes, with synthesis by diguanylate cyclases containing a GGDEF domain and degradation by phosphodiesterases containing either an HD-GYP or an EAL domain. These enzymes are ubiquitous in bacteria with up to 50 encoded in some genomes, the specific functions of which are mostly unknown.

**Results:** We used comparative analyses to identify genomic patterns among genes encoding proteins with GGDEF, EAL, and HD-GYP domains in five orders of the class Alphaproteobacteria. GGDEF-containing sequences and GGDEF-EAL hybrids were the most abundant and had the highest diversity of co-occurring auxiliary domains while EAL and HD-GYP containing sequences were less abundant and less diverse with respect to auxiliary domains. There were striking patterns in the chromosomal localizations of the genes found in two of the orders. The Rhodobacterales' EAL-encoding genes and Rhizobiales' GGDEF-EAL-encoding genes showed opposing patterns of distribution compared to the GGDEF-encoding genes. In the Rhodobacterales, the GGDEF-encoding genes showed a tri-modal distribution with peaks mid-way between the origin (*ori*) and terminus (*ter*) of replication and at *ter* while the EAL-encoding genes peaked near *ori*. The patterns were more complex in the Rhizobiales, but the GGDEF-encoding genes were biased for localization near *ter*.

**Conclusions:** The observed patterns in the chromosomal localizations of these genes suggest a coupling of synthesis and hydrolysis of c-di-GMP with the cell cycle. Moreover, the higher proportions and diversities of auxiliary domains associated with GGDEF domains and GGDEF-EAL hybrids compared to EAL or HD-GYP domains could indicate that more stimuli affect synthesis compared to hydrolysis of c-di-GMP.

**Keywords:** Phosphodiesterase, Diguanylate cyclase, DNA replication, Cell cycle, HD-GYP, GGDEF, EAL

\*Correspondence: aslang@mun.ca

Department of Biology, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada

## Introduction

Bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) is a second messenger that was first described for its role in regulating cellulose biosynthesis in *Gluconacetobacter*



*xylinus* [1, 2], but which is now recognized as near-ubiquitous and affecting a large variety of processes in bacteria [3, 4]. Cellular concentrations of c-di-GMP are regulated in response to internal and external stimuli, and the resulting changes can be part of bacterial adaptation to changes in their environment [5]. The cellular levels of c-di-GMP are controlled by two groups of enzymes with opposing activities, where it is synthesized by diguanylate cyclases (DGCs) and degraded by c-di-GMP-specific phosphodiesterases (PDEs). DGCs have conserved GGDEF domains and synthesize c-di-GMP from two molecules of guanosine triphosphate (GTP) [6]. There are two distinct types of PDEs, with either EAL or HD-GYP domains, that degrade c-di-GMP. Both types are able to break c-di-GMP into the linear 5'-phosphoguanlyl (3'-5') guanosine (pGpG) dinucleotide [7, 8] which is then further broken down to two molecules of guanosine monophosphate (GMP) by an oligo-ribonuclease [9, 10]. PDEs of the HD-GYP type can also break c-di-GMP into two GMPs in one step [11, 12]. In addition, there are hybrid proteins that have both GGDEF and EAL domains and thus represent a “biochemical conundrum”. It has been suggested that these proteins can switch between synthesis and hydrolysis of c-di-GMP [13], with a protein's activity controlled by, for example, phosphorylation status [14] or dimerization [15]. However, it is also possible that one of the domains is not enzymatically functional. Based on the proteins characterized in detail, the most common scenarios are that only the EAL domain is functional or both domains are functional [16].

C-di-GMP levels can be controlled via transcriptional and translational regulation of gene expression, or through post-translational modification of the synthesis and degradation enzymes as a quicker response. Auxiliary domains can be present on the enzymes and include sensory, signalling and protein binding domains, and these can allow for rapid adaptation [17]. Cellular changes in c-di-GMP concentration can result from a variety of input and output signals that are detected by the enzymes or their regulators and that affect the production or degradation of c-di-GMP [18]. An analysis of genomic sequences from different bacterial phyla found that members of the phylum Proteobacteria encode the highest numbers of c-di-GMP-modulating enzymes [18].

In the Alphaproteobacteria, c-di-GMP has been examined for its role in many different processes, such as the symbiosis of *Sinorhizobium meliloti* with plant roots [19] and related to its effects on the regulatory network associated with the transcriptional regulator CtrA [20], which is highly conserved in this class [21]. The CtrA phosphorelay consists of the histidine kinase CckA, the phosphotransferase ChpT and the transcriptional regulator CtrA [22]. It has been suggested that its ancestral role in

alphaproteobacteria was related to the control of motility and recombination [23, 24], but there has also been work establishing a link between this phosphorelay and c-di-GMP with respect to regulation of the cell cycle and cell differentiation in *Caulobacter crescentus* [20, 25] and gene transfer agent (GTA) production in *Rhodobacter capsulatus* and *Dinoroseobacter shibae* [26, 27]. C-di-GMP affects the CtrA phosphorelay directly through effects on the enzymatic activity of CckA, which changes the phosphorylation level of CtrA and thus its activity [28, 29]. The concentration of c-di-GMP also appears to be affected by CtrA because loss of CtrA results in changes in the transcript levels of genes encoding c-di-GMP-metabolizing enzymes [30].

The chromosomal positioning of genes can affect their functions in different ways and have effects on multiple cellular processes. For example, gene location can influence the spatial distribution of proteins within cells due to transcription-coupled translation [31]. Positioning can also have effects with respect to the cell cycle because genes that are close to the origin of replication (*ori*) are replicated earlier and are therefore temporarily present in higher copies than genes that are closer to the terminus of replication (*ter*) [32]. An example where this has important implications was found in *Bacillus subtilis*, where it was shown that the temporal copy number imbalances due to the opposite localization of genes encoding members of a regulatory network influenced its output [33, 34]. Additionally, gene location can influence expression due to the state of DNA methylation through the cell cycle. The partially replicated portions of the chromosome are hemi-methylated during replication starting at *ori*, and methylation status can affect regulatory protein binding and transcription [35]. For example, and directly related to regulatory systems already discussed above, one of the promoters where transcription of *ctrA* initiates is only activated in the hemi-methylated state in *C. crescentus* [36]. It seems likely there are additional and broader implications of gene location related to CtrA because a previous analysis also showed that numerous genes that are connected to CtrA have conserved chromosome positions in members of the Alphaproteobacteria [37].

C-di-GMP-modulating enzymes are broadly distributed in phylogenetically and metabolically diverse bacteria. They are also very diverse with respect to their roles and regulation, with a wide range of stimuli affecting c-di-GMP levels, and only a small proportion of the total diversity of these enzymes has been characterized in detail [38]. Therefore, we were interested in identifying any underlying genomic properties that these enzymes might share. We performed a comparative analysis of sequences containing GGDEF, EAL, and HD-GYP

domains from five orders of the Alphaproteobacteria, the Rhodospirillales, Sphingomonadales, Rhodobacterales, Rhizobiales and Caulobacterales. We identified the auxiliary domains present with these c-di-GMP-metabolizing domains and attempted to identify patterns regarding enzyme occurrences, distributions, and chromosomal localizations.

## Methods

### Dataset

Protein sequences with identified EAL (PF00563), GGDEF (PF00990) or HD (PF01966) domains from genomes of bacteria within five orders of the Alphaproteobacteria (Rhodospirillales, Sphingomonadales, Rhodobacterales, Rhizobiales and Caulobacterales) were downloaded from the EMBL website on 6 August 2020 (GGDEF: <http://pfam.xfam.org/family/PF00990#tabview=tab7>; EAL: <http://pfam.xfam.org/family/PF00563#tabview=tab7>; HD: <http://pfam.xfam.org/family/PF01966#tabview=tab7>) [39]. Proteins with the HHExxxxxGYP motif from within the HD sequences were then selected and considered PDEs while the remaining HD sequences were considered auxiliary domains if they co-occurred with a c-di-GMP-metabolizing domain. Proteins with both EAL and GGDEF domains were placed in their own group (GGDEF\_EAL).

All analyses were done in R version 4.0.3 with the appropriate packages as needed (Table S1).

### Organism, domain, and genomic annotations

Sequence identifiers were extracted from the EMBL fasta files and used to access the respective organism information from UniProt (e.g., <https://www.uniprot.org/uniprot/V4RSF5.txt>) or EBI (e.g., <https://www.ebi.ac.uk/ena/browser/api/summary/PNQ76602>) [40]. The identifiers were also used to withdraw the domain information from Pfam (e.g., <http://pfam.xfam.org/protein/A0A0N0K049#tabview=tab0>). Domain annotations could not be withdrawn for all sequences due to inconsistent html path formatting, which reduced the dataset (Table 1). The identifiers were also used to obtain the NCBI protein identifiers from UniProt (e.g., <https://www.uniprot.org/uniprot/A0A0N0K3V8.txt>). Due to inconsistencies some sequences have different version numbers (e.g., <https://www.uniprot.org/uniprot/A0A0N0K3V8.txt?version=1>) and only version 1 was subsequently considered in such cases. All sequence identifiers and html paths can be found in Table S2. The NCBI identifiers were used to obtain genomic information from the gff and fasta files, downloaded from NCBI in GenBank format.

**Table 1** Genomes, genera, and species/strains available for analyses

Order	Closed genomes	One unambiguously identified <i>ori</i>	C-di-GMP-metabolizing domains	By genera <sup>a</sup>	By species or strains
Rhodospirillales	132	75	EAL	42	62
			GGDEF	47	71
			GGDEF_EAL	48	74
			HD-GYP	23	35
Sphingomonadales	27	17	EAL	22	145
			GGDEF	25	172
			GGDEF_EAL	26	174
			HD-GYP	8	18
Rhodobacterales	187	69	EAL	123	308
			GGDEF	132	333
			GGDEF_EAL	121	278
			HD-GYP	11	16
Rhizobiales	424	133	EAL	77	227
			GGDEF	90	257
			GGDEF_EAL	93	266
			HD-GYP	38	105
Caulobacterales	44	8	EAL	5	20
			GGDEF	5	28
			GGDEF_EAL	5	28
			HD-GYP	1	1

<sup>a</sup> Genera numbers include any Sphingomonadales sp., Rhodobacteraceae sp., Rhizobiales sp., and Caulobacteraceae sp. designations as one each

### Identification of chromosomal origins of replication

The origin of replication (*ori*) was identified for each chromosome using Ori-Finder and default settings [41]. The ptt files were generated (<https://github.com/sgivan/gb2ptt#gb2ptt>) from gbff files, downloaded from NCBI on 23 April 2019. Only chromosomes with one unambiguously identified *ori* were subsequently included in the investigation, which reduced the dataset (Table 1). The terminus of replication (*ter*) was assumed to be opposite *ori* on the circular chromosomes [42].

### Phylogenetic analysis

RpoB sequences (PF05000, RNA polymerase Rpb1, domain 4) were downloaded for the members of each order and their NCBI identifiers were determined. Alignments were done using MAFFT with L-INS-i option [43] in Geneious version 11.0.5 [44]. Phylogenetic trees were reconstructed using IQ-TREE version 2.1.4 [45], with the best substitution matrix identified using ModelFinder. The robustness of the analysis was tested using a bootstrap test (1000 replicates) [46] and a hill-climbing nearest-neighbor interchange search [45, 47]. Trees were modified and annotated in iTOL version 5 [48].

## Results

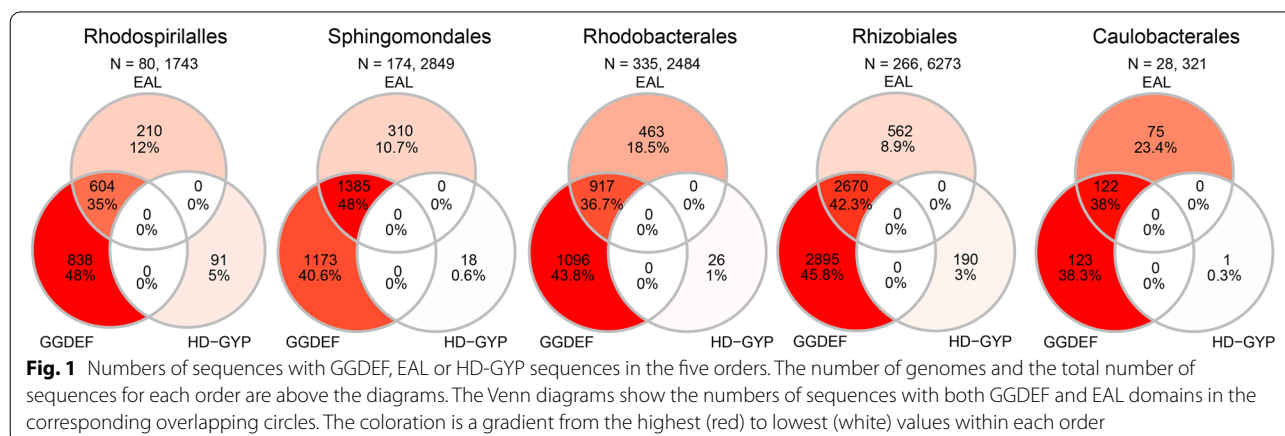
### Occurrence of c-di-GMP-modulating domains

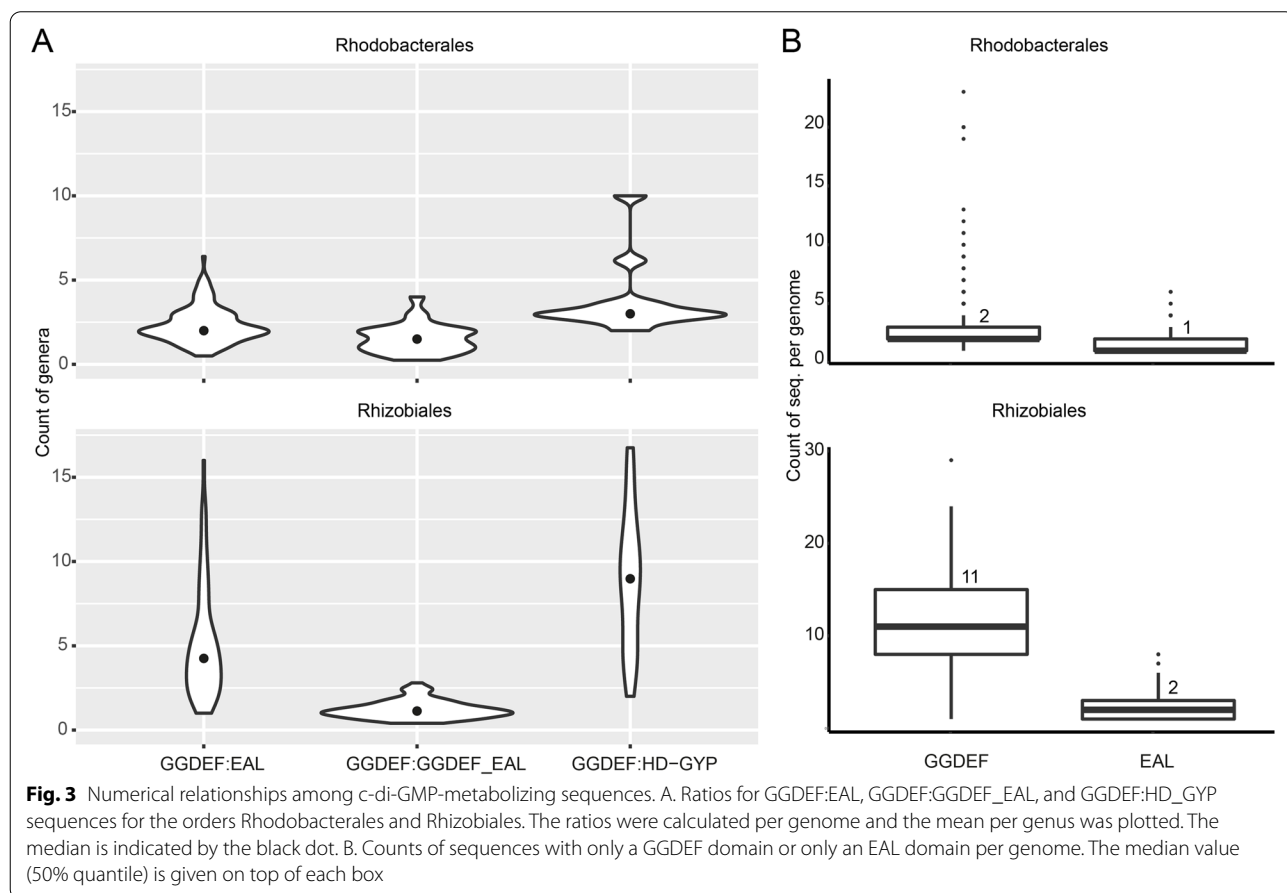
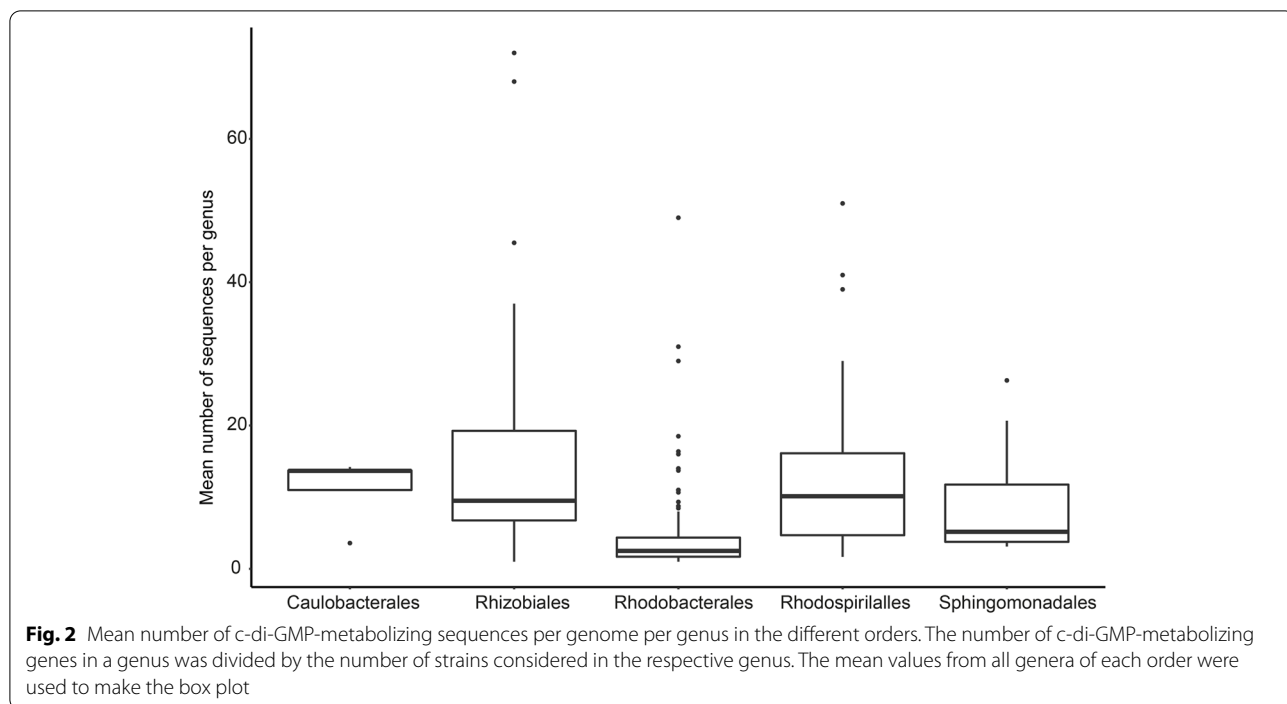
We quantified the genes encoding the domains associated with c-di-GMP synthesis and degradation in members of the five alphaproteobacterial orders. This included those that contained one of the GGDEF, EAL, or HD-GYP domains or both GGDEF and EAL domains. The GGDEF and GGDEF\_EAL sequences accounted for the highest proportions in all five orders at 35–48%, followed by proteins containing an EAL domain that ranged between 8.9% and 23.4% of all sequences (Fig. 1). The HD-GYP domain-containing sequences made up the smallest share, accounting for only 0.3–5% of all sequences, and

co-occurrence of GGDEF or EAL with an HD-GYP domain was not observed (Fig. 1). Each c-di-GMP-metabolizing domain was found almost exclusively once per sequence, but there were a few exceptions (Table S4).

Next, the numbers of c-di-GMP-metabolizing sequences in different genera were compared by calculating the mean number of sequences per genus (Fig. 2, Table S3). The c-di-GMP-metabolizing sequences per genus decreased from the Rhizobiales, Rhodospirillales, Caulobacterales, Sphingomonadales to the Rhodobacteriales, but ranges of 1–72 (*Rhodomicrobium* and *Neorhizobium*), 1.7–51 (*Ferruginivarius* and *Thalassospira*), 3.6–14.2 (*Phenylobacterium* and *Caulobacter*), 3–25.4 (*Croceicoccus* and *Novosphingobium*), and 1–49 (*Salicibibacter* and *Roseibium*) were observed in the respective individual orders.

For the subsequent investigation of the numerical relationships among the various domains, all orders were analyzed (Figure S1), but due to the larger number of available sequences and therefore more unambiguous results, we focused in particular on the Rhodobacteriales and Rhizobiales. Examination of the per genus ratios of genes encoding synthesizing enzymes to those encoding hydrolyzing enzymes, i.e., GGDEF:EAL, revealed that this ratio was always 2 or higher (Fig. 3A). However, the ratio was more consistently close to 2 across the Rhodobacteriales (0.5–6) as compared to the Rhizobiales (1–16), where there were more frequently higher numbers of GGDEF sequences and more variation among members of this order. When the numbers of GGDEF and EAL domain sequences per genome were examined (Fig. 3B), we found that the medians were 2 and 11 for GGDEF sequences and 1 and 2 for EAL sequences in the Rhodobacteriales and Rhizobiales, respectively. This again shows that genes encoding the synthesizing enzymes occur more frequently than those encoding hydrolyzing enzymes in both orders. The GGDEF:GGDEF\_EAL







ratios peaked at 1 in the studied orders except the Rhodobacterales where more variability was observed and a higher proportion of members showed higher ratios (Fig. 3A, Figure S2). Interestingly, the relationships of the GGDEF:EAL and GGDEF:GGDEF\_EAL ratios showed opposite patterns in the Rhodobacterales and Rhizobiales. While the GGDEF:EAL ratios were less variable and most consistently at 2 in the Rhodobacterales, there was much greater variability in the Rhizobiales. Conversely, there was more variability in the GGDEF:GGDEF\_EAL ratios in the Rhodobacterales but a distinct peak at 1 in the Rhizobiales. The relationship of GGDEF:HD-GYP domains was found to be fairly consistent at 2.5:1 in the Rhodobacterales but highly variable in the Rhizobiales (Fig. 3A).

The large variability in numbers of c-di-GMP-metabolizing proteins among organisms stimulated us to investigate their evolutionary relationships. Therefore, the number of c-di-GMP enzymes present in different species was evaluated in a phylogenetic context (Figure S4). Some closely related groups were found in which the numbers of c-di-GMP genes were similar. In the Rhizobiales there was a large cluster in which the c-di-GMP-metabolizing gene numbers were elevated, and which consisted of several genera, including *Devosia*, *Fulvimarina* and *Rhizobium*. Smaller additional clusters with increased c-di-GMP numbers that were less closely related were also observed. In the Rhodobacterales, the closely related genera *Stapia* and *Labrenzia* stood out with their high c-di-GMP-metabolizing gene numbers. A connection between phylogeny and c-di-GMP-metabolizing gene number could also be observed in the Rhodospirillales. Here there were three clusters of organisms that had increased gene numbers and one notable group was made up of three genera including *Magnetosprillum*, *Magnetovibrio* and *Telmatosprillum*. A clear connection between phylogenetic relationships and numbers of c-di-GMP-metabolizing genes was not observed in the Sphingomonadales, and it is difficult to make any statement for the Caulobacterales because of the lower genome and gene numbers.

#### Relationship between gene numbers, genome size, and location of c-di-GMP-metabolizing genes on secondary chromosomes

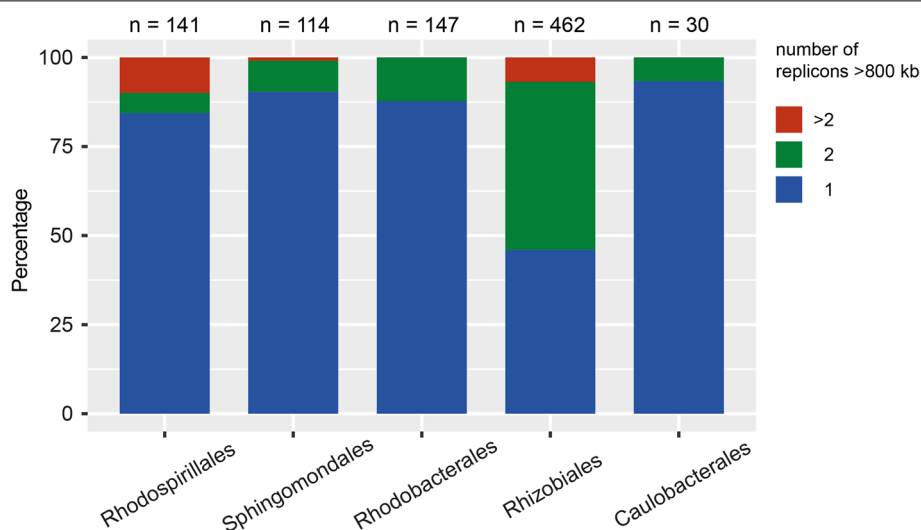
There was a statistically significant positive correlation between chromosome size and the number of c-di-GMP-metabolizing genes in all five orders (Figure S5). We only included the largest replicon in this analysis, although c-di-GMP-metabolizing genes were also found on secondary chromosomes and extrachromosomal replicons. In five genomes from different genera of the Rhodospirillales, six genomes from three genera in the

Sphingomonadales, five genomes from five different genera of the Rhodobacterales, 23 genomes from 13 genera of the Rhizobiales, and one genome of the Caulobacterales c-di-GMP-metabolizing genes were found outside of the largest replicon (Table S5). In *Nitrospirillum amazonense* CBAmc (Rhodospirillales), *Rhizobium* sp. NXC24 (Rhizobiales) and *Asticcacaulis excentricus* CB 48 (Caulobacterales) more c-di-GMP genes were found on the second-largest replicon and in *Paracoccus denitrificans* PD1222 (Rhodobacterales) equal numbers of c-di-GMP-metabolizing genes were found on the largest and second-largest replicons.

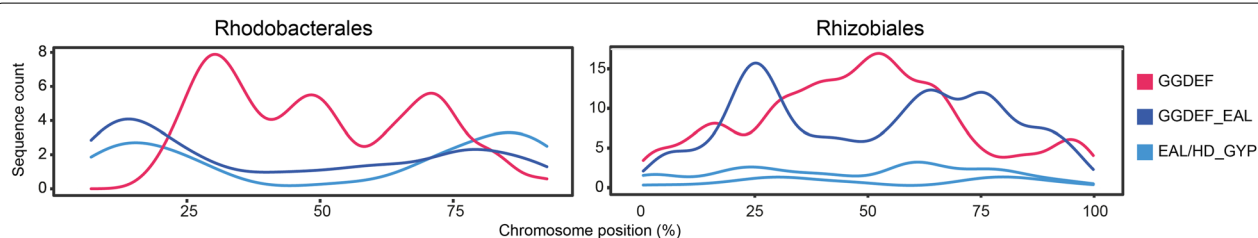
Secondary chromosomes (defined as replicons > 800 kb that are not the largest replicons in the genome) contain genes that evolve faster [49] and are more common in the Rhizobiales (Fig. 4). We investigated if c-di-GMP-metabolizing genes were found outside of the largest chromosome more often when secondary chromosomes were present. We found that only a small fraction of the genomes examined in this study had secondary chromosomes in four of the orders (14.9% or 21 genomes of the Rhodospirillales, 8.8% or 10 genomes of the Sphingomonadales, 10% or 15 genomes of the Rhodobacterales, and 6.7% or 2 genomes of the Caulobacterales) whereas this was higher for the Rhizobiales (44% or 204 genomes). There were c-di-GMP-metabolizing genes on the secondary chromosomes in all orders and these accounted for 21.3, 21.1, 30, 31.7 and 73.3% of all c-di-GMP-metabolizing genes in the Rhodospirillales, Sphingomonadales, Rhodobacterales, Rhizobiales, and Caulobacterales, respectively. We note that the high percentage of c-di-GMP-metabolizing genes identified on secondary chromosomes in the Caulobacterales is based on only two genomes. Overall, the results indicate that the presence of secondary chromosomes did not result in a greater proportion of c-di-GMP-metabolizing genes located there.

#### Chromosomal organization patterns of c-di-GMP-metabolizing genes

As discussed above, location on the chromosome can affect gene expression. We therefore wanted to examine the localization of c-di-GMP-metabolizing genes on chromosomes relative to the origin (*ori*) and terminus (*ter*) of replication. No obvious trend was observed in the Rhodospirillales, while GGDEF and GGDEF\_EAL sequences seemed less prevalent near *ter* in the Sphingomonadales (Figure S6). The number of genes included in the analysis for the Sphingomonadales EAL group and all groups for the Caulobacterales were so low that patterns might not be obvious even if present. However, interesting patterns were evident in the Rhizobiales and Rhodobacterales (Fig. 5). In the Rhodobacterales the EAL and



**Fig. 4** Proportions of genomes with one, two or more than two replicons >800 kb in the five orders. The total numbers of genomes in each order are above the plot



**Fig. 5** Chromosomal locations of *c*-di-GMP-metabolizing genes. Cumulative distributions of *c*-di-GMP-metabolizing genes on the chromosomes of Rhodobacterales and Rhizobiales, with lengths normalized to 100% where *ori* is at 0% and *ter* is at 50%. The color-coded lines represent the estimate of the kernel density. Only closed genomes with one unambiguously determined *ori* were used in this analysis

GGDEF\_EAL sequences were predominately found near *ori* whereas GGDEF sequences were predominately not close to *ori* and showed a tri-modal distribution with peaks mid-way between *ori* and *ter* and around *ter*. In the Rhizobiales, clear patterns were observed for the GGDEF and GGDEF\_EAL sequences, which both showed multiple peaks but with opposing patterns. The distribution of the GGDEF sequences showed three peaks, with the largest near *ter* and two smaller peaks near *ori*. The GGDEF\_EAL sequences peaked where the GGDEF sequences were lowest, mid-way between *ori* and *ter*. Although there were far fewer sequences, the Rhizobiales HD-GYP group showed a similar trend as the GGDEF\_EAL sequences, while there was no obvious pattern for the EAL sequences.

Comparison of the similarities of distributions among the groups of genes indicated that the Rhodobacterales EAL and GGDEF\_EAL genes were similarly distributed (two-sample Kolmogorov–Smirnov test;  $p$ -value=0.16) while the EAL and GGDEF as well as the GGDEF and

GGDEF\_EAL pairs were distributed differently (two-sample Kolmogorov–Smirnov test;  $p$ -values=0.009 and 0.0008, respectively). The Rhizobiales GGDEF and GGDEF\_EAL genes were also distributed differently (two-sample Kolmogorov–Smirnov test;  $p$ -value=0.04).

#### Additional domains on *c*-di-GMP-metabolizing proteins

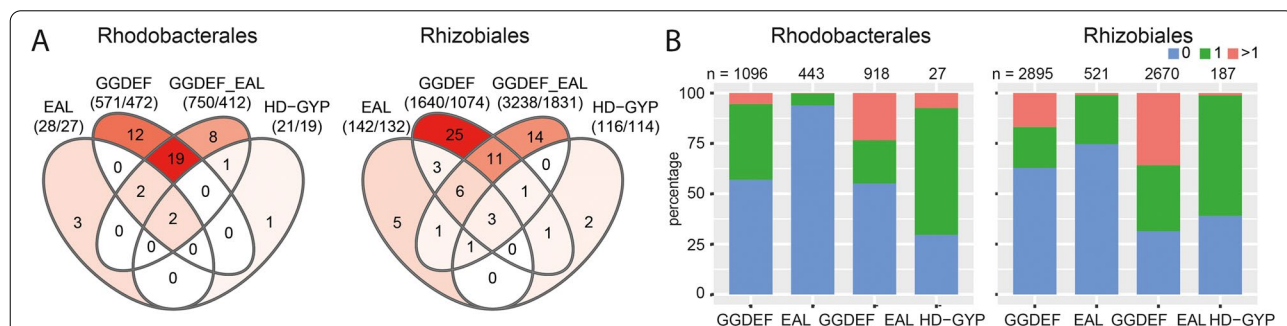
It has previously been documented that proteins with *c*-di-GMP-metabolizing domains frequently contain additional domains [20], hereafter referred to as auxiliary domains, which presumably function in many cases to regulate the *c*-di-GMP-related enzymatic activities. Only the Rhodobacterales and Rhizobiales are discussed in detail here because of the larger numbers of sequences available for these orders, but similar trends were also observed in the other three (Table S6, Figure S7). Auxiliary domains were associated with all four *c*-di-GMP sequence groups and there were 101 different auxiliary domains found across all five orders and sequence groups. We note that the auxiliary domains analyzed

here are those that are identified and specified in databases but recognize that some of the sequences will have uncharacterized domains that are not captured there. We plotted the length of EAL-containing sequences and this showed that all those with identified auxiliary domains were >375 amino acids long (Figure S8). The proportions of those without identified auxiliary domains that were <375 amino acids long were 49% in the Rhizobiales and 82% in the Rhodobacterales, indicating that some of these proteins likely contain auxiliary domains but these remain to be recognized and annotated in the sequence databases. The same analysis with GGDEF sequences revealed that all sequences containing identified auxiliary domains were >275 amino acids long (Figure S8). The proportions of those without identified auxiliary domains that were <275 amino acids long were 13% in the Rhizobiales and 30% in the Rhodobacterales and, therefore, most of these sequences likely also contain currently unannotated auxiliary domains.

In both the Rhodobacterales and Rhizobiales the GGDEF group had the highest variability among auxiliary domains, followed by GGDEF\_EAL, EAL and HD-GYP sequences (Fig. 6A). However, this could be driven by the higher number of sequences containing GGDEF domains compared to other domains (Fig. 1). The GGDEF and GGDEF\_EAL groups had the greatest overlap of auxiliary domains whereas there were only a few unique domains present with the EAL and HD-GYP domain-containing sequences. Overall, there were uniform distributions of sequences that contain none, one, or more than one auxiliary domain (Fig. 6B, Figure S7). The HD-GYP group had the highest proportion of sequences with auxiliary domains, followed by the GGDEF\_EAL, GGDEF and EAL groups (Fig. 6B). The GGDEF\_EAL group had the

biggest proportion of sequences that had more than one auxiliary domain on individual proteins.

Some auxiliary domains were more commonly found in certain groups and some of these co-occurrences were conserved across the five orders (Tables S6 and S7). A previous study reported that cGMP-specific phosphodiesterases, adenylyl cyclases and FhlA (GAF) and Per-Arnt-Sim (PAS) were the most common auxiliary domains associated with GGDEF domains in various bacterial species [17]. The GAF domain is a sensory domain involved in light sensing and it and the PAS domain have been found in phytochromes [17, 50]. In our GGDEF sequences, the response regulator receiver (REC) domain and PAS domain variants dominated. Cognate histidine kinases modulate REC domain-containing proteins through their phosphorylation status via their kinase and phosphatase activities, which are themselves regulated by various signals. The phosphorylation status of the REC domain then controls the activity of the associated output domain (e.g., GGDEF). In the EAL group REC domains, CSS-motif (Pfam PF12792) domains and GAF\_2 domains were most common. CSS-motif domains are known for roles in redox sensing [17]. The Caulobacterales EAL sequences were an exception, because these were most often associated with histidine kinase and phosphotransferase domains that act upstream of REC domains in histidyl-aspartyl phosphorelay systems. In the GGDEF\_EAL sequences, the PAS subfamilies PAS\_3, PAS\_4, PAS\_7 and PAS\_9, as well as the MHYT domain were most common. The MHYT domain consists of six transmembrane segments and it has been suggested to function in O<sub>2</sub>, NO and CO sensing [51]. In the HD-GYP sequences HD\_5 and two domains of unknown function, DUF3369 and DUF3391, were the most prevalent.

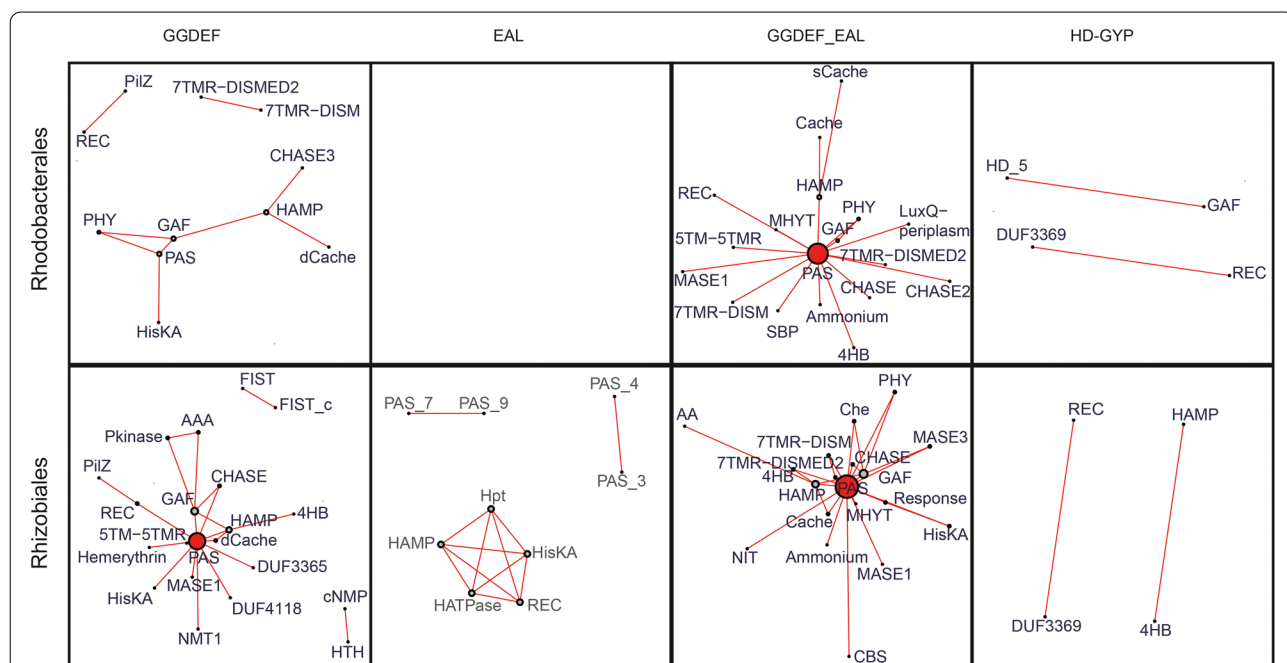


**Fig. 6** Occurrence of auxiliary domains on c-di-GMP-metabolizing proteins of the different enzyme groups. A. Numbers of different auxiliary domains that can be found for each group and shared among groups. The first number below the group identification (EAL, GGDEF, GGDEF\_EAL, HD-GYP) indicates the number of auxiliary domains in the respective group and the second number indicates the number of sequences these domains are found in. The c-di-GMP-metabolizing domains themselves are not included in this analysis. The color code of the Venn diagram represents the domain counts from the highest (red) to zero (white). B. Percentage of sequences with none, one, or more than one auxiliary domain. The number of sequences included in this analysis is given above the group identification. Repeated occurrence of a domain in a sequence was counted as one



Despite detailed knowledge on the structure and function of DGCs and PDEs, it has remained challenging to assign physiological roles to individual proteins. Analysis of the co-occurrence of additional domains might aid in assigning those roles. Therefore, we next investigated which additional domains occurred together and constructed co-occurrence networks (Fig. 6, Table S8). We focused on the Rhodobacterales and Rhizobiales because more sequences with more than one auxiliary domain were available for these orders. Most of the Rhodobacterales GGDEF sequences that had more than one auxiliary domain had co-occurrences of two specific auxiliary domains (Fig. 7). Exceptions were phytochrome (PHY), PAS, GAF, histidine kinase, adenylate cyclase, methyl-accepting protein and phosphatase (HAMP) domains, which co-occurred with two or three other domains. PAS domains were dominant in co-occurrences with many other domains in the GGDEF and GGDEF\_EAL groups of both orders as well as the Rhizobiales' EAL group (Fig. 7). Linkage of one domain with a variety of others creates complex patterns, such as found for the GGDEF sequences of both orders where calcium channels and chemotaxis receptors (dCache\_1), GAF\_2, HAMP and cyclase/histidine kinase-associated sensory extracellular (CHASE3) domains formed a network.

The Cache and CHASE domains are extra-cytoplasmic sensory domains [52, 53] while the HAMP domain is usually found in integral membrane proteins that transmit conformational changes from periplasmic ligand-binding domains to cytoplasmic domains as part of histidyl-aspartyl phosphorelay signaling [54]. In the GGDEF\_EAL sequences of both orders and the GGDEF sequences of the Rhizobiales, the PAS domains were notable because they are the domains connected with the most other auxiliary domains. Interestingly, the EAL sequences of the Rhizobiales had one cluster composed of the same domains that are most prevalent in the EAL sequences of the Caulobacterales (Table S7). These are the HisKA domain (activated via dimerization and able to transfer a phosphoryl group often as part of histidyl-aspartyl phosphorelay systems [55]), the Hpt domain that mediates phosphotransfer in histidyl-aspartyl phosphorelay systems [56], the HAMP domain, and HATPase that is found in multiple ATPases such as histidine kinases [57]. This shows that the EAL sequences, when linked to auxiliary domains, are often part of signaling cascades, especially in the Rhizobiales and Caulobacterales. The HD-GYP sequences showed two connections per order, one of which seemed to be conserved in the Rhodobacterales and Rhizobiales and consisted of the DUF3369 and REC domains.



**Fig. 7** Weighted graphs representing the co-occurrences of auxiliary domains with c-di-GMP-metabolizing sequences. Auxiliary domains occurring together are connected by lines with the size and red color of the node indicating higher frequency of co-occurrence with other domains. Lengths of edges represent the number of times the connected domains co-occur, and the sizes of the points indicate the number of times these domains occur. All full domain names are provided in Table S8

## Discussion

### Associations with diverse auxiliary domains suggest a wide variety of signals affect DGC activity

Our analysis of the occurrence of the EAL, GGDEF, GGDEF\_EAL and HD-GYP sequences in orders of the Alphaproteobacteria showed that the GGDEF and GGDEF\_EAL domains made up the biggest proportions in all orders, followed by the EAL domains, while the HD-GYP domains accounted for the smallest share. Compared to results from a study on c-di-GMP-metabolizing gene distributions among prokaryotes, which found the overall proportions to be 50.4% GGDEF, 16.1% EAL and 33.5% GGDEF\_EAL [11], the alphaproteobacterial orders have slightly lower GGDEF and higher GGDEF\_EAL proportions. Moreover, the GGDEF and GGDEF\_EAL sequences are associated with more different types of auxiliary domains and have a proportionally higher occurrence of auxiliary domains, respectively. This suggests that the GGDEF\_EAL proteins more frequently respond to signals/stimuli, but the GGDEF-only proteins integrate a broader variety of signals. Thus, since GGDEF domain sequences are more abundant and seem to have more and more diverse auxiliary domains than PDE domain sequences, it could be that the synthesis of c-di-GMP is mainly controlled in response to extracellular and intracellular signals while its degradation is more unspecific. Since the GGDEF\_EAL sequences of the Rhizobiales, like the EAL sequences of the Rhodobacterales, show a lower diversity of auxiliary domains, they too could be responsible for unspecific degradation while increases in c-di-GMP are more regulated. However, we note that this analysis is limited by its reliance on detecting recognized auxiliary domains while it is likely that some of these proteins contain currently unrecognized auxiliary domains.

### Importance of EAL-type PDE domains in Proteobacteria

Proteins with only EAL domains outnumbered those with HD-GYP domains at least two-fold in all orders. This agrees with a previous analysis of these domains in several phyla where the Proteobacteria, with the exception of the Deltaproteobacteria, and Oligoflexia were the only investigated phyla in which EAL domains outnumbered HD-GYP domains [58]. The driving forces behind the trends for relative abundances of these two different types of PDEs are not clear and likely require a larger phylogenetic analysis to untangle. More information on the specific roles of individual proteins is also required. The possible activities of proteins with both GGDEF and EAL domains, which are even more abundant than PDEs without GGDEF domains, further complicates the situation.

### Shared genomic features of the Rhizobiales GGDEF\_EAL and Rhodobacterales EAL sequences

Interestingly, multiple commonalities exist between the GGDEF\_EAL sequences of the Rhizobiales and the EAL sequences of the Rhodobacterales. Both gene groups are biased for localization away from *ter*, and their relative abundances compared to GGDEF sequences are reversed in the two orders. The GGDEF:EAL ratio is very consistent in the Rhodobacterales but there is no such consistency in the Rhizobiales. Conversely, while the GGDEF:GGDEF\_EAL ratios were more varied in the Rhodobacterales, they were much more consistent in the Rhizobiales. This could indicate that the roles of the EAL sequences in the Rhodobacterales are swapped with GGDEF\_EAL sequences in the Rhizobiales. However, the hybrid nature of GGDEF\_EAL sequences makes this difficult to conclude. The two activities can be switched, e.g., by dimerization, which is required for GGDEF but not for EAL activity [15], or through regulation by auxiliary domains [14, 59, 60]. However, a study of the conservation of amino acid patterns showed that the catalytic activity in hybrid sequences is most often preserved in both domains or only in the EAL domain [16]. Future studies must show whether the Rhizobiales hybrid sequences have mainly retained EAL activity and thereby compensate for the lack of EAL sequences near *ori*, assuming they are involved in the same functions as the Rhodobacterales EAL sequences that are also positioned near *ori*. This could potentially be initially evaluated through a large-scale bioinformatic analysis of the enzymatic domains in the Rhizobiales GGDEF\_EAL hybrids to look for conservation of known critical residues required for DGC and PDE activity.

### Conserved chromosomal positioning

In the two orders with the most available data, the Rhodobacterales and Rhizobiales, there is a clear conservation of the Rhodobacterales EAL- and GGDEF\_EAL- and the Rhizobiales GGDEF\_EAL-encoding genes away from *ter* while the GGDEF-encoding genes are predominant on the *ter*-proximate half of the chromosome in both orders. Overall, the concentrations of GGDEF genes peak when the EAL and GGDEF\_EAL genes in the Rhodobacterales and the GGDEF\_EAL genes in the Rhizobiales drop. This could indicate that there is more c-di-GMP degradation in the cell near the *ori* and more synthesis near the *ter* in the Rhodobacterales, which could also apply to the Rhizobiales should it turn out that the hybrid sequences primarily act as PDEs (discussed above).

There are multiple potential effects caused by the chromosomal locations of specific genes. The observed chromosomal localization patterns revealed in this study

might affect cellular c-di-GMP concentrations during the cell cycle. Genes that are close to *ori* are replicated earlier than genes that are close to *ter*, which leads to a temporary copy number imbalance between genes at these two locations [32]. In *B. subtilis*, the opposite location of two genes encoding components of a phosphorylation relay with respect to *ori* and *ter* leads to temporal copy number imbalances, and this allows spore formation to only take place at the end of the cell cycle when the balance between the regulators is restored [33, 34]. In *Vibrio cholerae*, moving genes from *ori* to *ter* and thus reducing their copy number during the cell cycle has an impact on growth and infectivity [61, 62]. Such copy number imbalances can be pronounced in organisms that initiate multiple rounds of DNA replication within individual cells, such as *Escherichia coli* [63], although there is no evidence this occurs in members of the alphaproteobacteria. Regardless, it is possible that the biased localizations of genes encoding c-di-GMP-metabolizing enzymes we observed could have some effects on cellular c-di-GMP concentrations through temporary copy number imbalances, but future experimental work is required to evaluate this.

Another effect of localization could be manifested through DNA methylation, where the chromosomal DNA changes from fully methylated to hemi-methylated during replication. This change in methylation can affect gene transcription. For example, the p1 promoter of the *ctrA* gene in *C. crescentus* is only active in the hemi-methylated state [36]. Thus, *ctrA*, which is localized near *ori*, is transcribed more during DNA replication because it is hemi-methylated right at the beginning of the cycle. However, any broad role of methylation in regulating transcription of genes encoding c-di-GMP-metabolizing enzymes is currently unknown and future work is required to investigate this possibility.

## Conclusions

C-di-GMP-metabolizing enzymes are very diverse, and the specific roles and functions of only a few of these proteins are known. In this study new patterns and common properties for these proteins were identified in members of the Alphaproteobacteria. We systematically examined gene occurrence, localization on the genome, and the presence of auxiliary domains. In the Rhodobacterales and Rhizobiales, the EAL and GGDEF\_EAL sequences, respectively, are primarily located away from *ter* while GGDEF sequences are biased towards *ter*. Additionally, the EAL and GGDEF\_EAL domain-containing sequences show lower diversity and occurrence of auxiliary domains compared to the GGDEF sequences. There are several known examples in which chromosome localization of genes is important, and this can manifest in different

ways such as through changes in copy number and methylation status during the cell cycle. The patterns we found support the suggestion that the chromosomal localization of c-di-GMP-metabolizing genes is important in these bacteria. Our findings also support the notion that the synthesis of c-di-GMP is more regulated and responsive to a variety of specific signals whereas its degradation might be less regulated and dependent on different stimuli.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-09072-9>.

**Additional file 1: Figure S1.** Numerical relationships between the number of GGDEF and EAL (GGDEF:EAL) sequences by genera. The ratios were calculated per genome and the mean per genus was plotted. **Figure S2.** Numerical relationships between the number of GGDEF and GGDEF\_EAL (GGDEF:GGDEF\_EAL) sequences by genera. The ratios were calculated per genome and the mean per genus was plotted. **Figure S3.** Numerical relationships between the number of GGDEF and HD-GYP (GGDEF:HDGYP) sequences by genera. The ratios were calculated per genome and the mean per genus was plotted. **Figure S4.** Numbers of c-di-GMP sequences in a phylogenetic context. Phylogenetic relationships are based on RpoB sequences. All alignments were done using MAFFT with LINS-i option. Bootstrap values based on 1000 replicates and hill-climbing nearest-neighbor interchange search were used. A. Rhizobiales. B. Caulobacterales. C. Rhodobacterales. D. Rhodospirillales. E. Sphingomonadales. **Figure S5.** Relationships between chromosome size and the number of encoded c-di-GMP enzymatic domains. Spearman's rank correlation was used to evaluate the significance. Only the biggest replicon, considered the main chromosome, was included in this analysis. **Figure S6.** Chromosomal locations of c-di-GMP-associated genes. Cumulative distributions of c-di-GMP-associated genes on the chromosomes, with lengths normalized to 100% where *ori* is at 0% and 100% and *ter* is at 50%. The red line indicates the estimate of the kernel density. In this analysis only closed genomes with one unambiguously identified *ori* were used. **Figure S7.** Secondary domains that are present along with the different c-di-GMP-associated enzyme groups. A. Shared and individual secondary domains. The c-di-GMP-modulating domains are not included in this analysis. The color code of the Venn diagram represents domain counts from highest (red) to zero (white). B. Number of sequences that have zero, one, or more than one secondary domain. **Figure S8.** Relationships between protein length and presence of detected auxiliary domains. The sequences with EAL and GGDEF domains were segregated based on the occurrence of auxiliary domains. The minimal amino acid lengths for proteins containing auxiliary domains (left panel) were identified as 375 for EAL proteins and 275 for GGDEF proteins (blue dashed lines). This threshold was then used to calculate the percentage of sequences without identified auxiliary domains that were shorter and longer than these minimal lengths (right panel).

**Additional file 2: Table S1.** R packages used for analyses.

**Additional file 3: Table S2.** Compilation of Pfam, EBI and NCBI sequence identifiers.

**Additional file 4: Table S3.** Sequence counts by genera. The mean number of sequences is calculated per genus. Blue indicates a "genus" that is not included in the analysis because it does not represent an actual genus.

**Additional file 5: Table S4.** Domains found in all c-di-GMP-associated protein sequences. The cyclic di-GMP-associated domains are indicated in light blue. Values >0 for other domains are indicated in red.

**Additional file 6: Table S5.** Count of c-di-GMP-associated genes on chromosomes and plasmids. The primary chromosome is defined as the largest replicon of a genome. A "c" at the start of a column name indicates information regarding the chromosome while "p" indicates plasmids. Blue

coloration indicates those genomes that have c-di-GMP-associated genes on secondary chromosomes or extrachromosomal replicons.

**Additional file 7: Table S6.** Occurrence of secondary domains with cyclic di-GMP-modulating domain sequences.

**Additional file 8: Table S7.** Frequency of association of all secondary domains that co-occur with c-di-GMP-associated domains.

**Additional file 9: Table S8.** All domains found associated with one of the c-di-GMP-associated domains examined in this study with associated Pfam HTML paths and Pfam IDs.

## Acknowledgements

Not applicable.

## Authors' contributions

SK performed the analyses and original draft preparation of the manuscript. ASL performed supervision and writing, review, and editing of the manuscript. The authors read and approved the final manuscript.

## Funding

SK was partially supported by funding from the Memorial University of Newfoundland School of Graduate Studies. This research in ASL's lab was supported by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-04636 and RGPIN-2022-03791).

## Availability of data and materials

Data were withdrawn from the databases Pfam, EBI and NCBI. For identification hyperlinks or protein identifiers are supplied in the Supplementary Table S2 and in the Methods section.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests. The funders had no role in the design or performance of the study, or in the decision to publish the results.

Received: 19 May 2022 Accepted: 17 October 2022

Published online: 16 December 2022

## References

- Ross P, Weinhouse H, Aloni Y, Michaeli D, Weinberger-Ohana P, Mayer R, et al. Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature*. 1987;325:279–81. <https://doi.org/10.1038/325279a0>.
- Ross P, Aloni Y, Weinhouse H, Michaeli D, Weinberger-Ohana P, Mayer R, et al. Control of cellulose synthesis *Acetobacter xylinum*. A unique guanyl oligonucleotide is the immediate activator of the cellulose synthase. *Carbohydr Res*. 1986;149:101–17. [https://doi.org/10.1016/S0008-6215\(00\)90372-0](https://doi.org/10.1016/S0008-6215(00)90372-0).
- Valentini M, Filloux A. Multiple roles of c-di-GMP signaling in bacterial pathogenesis. *Annu Rev Microbiol*. 2019;73:387–406. <https://doi.org/10.1146/annurev-micro-020518-115555>.
- Römling U, Galperin MY, Gomelsky M. Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev*. 2013;77:1–52. <https://doi.org/10.1128/mmr.00043-12>.
- Krasteva PV, Sondermann H. Versatile modes of cellular regulation via cyclic dinucleotides. *Nat Chem Biol*. 2017;13:350–9. <https://doi.org/10.1038/nchembio.2337>.
- Schirmer T. C-di-GMP synthesis: structural aspects of evolution catalysis and regulation. *J Mol Biol*. 2016;428:3683–701. <https://doi.org/10.1016/j.jmb.2016.07.023>.
- Stelitano V, Giardina G, Paiardini A, Castiglione N, Cutruzzola F, Rinaldo S. C-di-GMP hydrolysis by *Pseudomonas aeruginosa* HD-GYP phosphodiesterases: analysis of the reaction mechanism and novel roles for pGpG. *PLoS ONE*. 2013;8:e74920. <https://doi.org/10.1371/journal.pone.0074920>.
- Christen M, Christen B, Folcher M, Schauerte A, Jenal U. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. *J Biol Chem*. 2005;280:30829–37. <https://doi.org/10.1074/jbc.M504429200>.
- Cohen D, Mechold U, Nevenzal H, Yarmiyhu Y, Randall TE, Bay DC, et al. Oligoribonuclease is a central feature of cyclic diguanylate signaling in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci*. 2015;112:11359–64. <https://doi.org/10.1073/pnas.1421450112>.
- Orr MW, Donaldson GP, Severin GB, Wang J, Sintim HO, Waters CM, et al. Oligoribonuclease is the primary degradative enzyme for pGpG in *Pseudomonas aeruginosa* that is required for cyclic-di-GMP turnover. *Proc Natl Acad Sci*. 2015;112:E5048–57. <https://doi.org/10.1073/pnas.1507245112>.
- Bellini D, Caly DL, McCarthy Y, Bumann M, An S-Q, Dow JM, et al. Crystal structure of an HD-GYP domain cyclic-di-GMP phosphodiesterase reveals an enzyme with a novel trinuclear catalytic iron centre. *Mol Microbiol*. 2014;91:26–38. <https://doi.org/10.1111/mmi.12447>.
- Ryan RP, Fouhy Y, Lucey JF, Crossman LC, Spiro S, He Y-W, et al. Cell–cell signaling in *Xanthomonas campestris* involves an HD-GYP domain protein that functions in cyclic di-GMP turnover. *Proc Natl Acad Sci*. 2006;103:6712–7. <https://doi.org/10.1073/pnas.0600345103>.
- Ryan RP, Fouhy Y, Lucey JF, Dow JM. Cyclic di-GMP signaling in bacteria: recent advances and new puzzles. *J Bacteriol*. 2006;188:8327–34. <https://doi.org/10.1128/JB.01079-06>.
- Levet-Paulo M, Lazzaroni J-C, Gilbert C, Atlan D, Doublet P, Vianney A. The atypical two-component sensor kinase Lpl0330 from *Legionella pneumophila* controls the bifunctional diguanylate cyclase-phosphodiesterase Lpl0329 to modulate bis-(3'-5')-cyclic dimeric GMP synthesis. *J Biol Chem*. 2011;286:31136–44. <https://doi.org/10.1074/jbc.M111.231340>.
- Jenal U, Malone J. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annu Rev Genet*. 2006;40:385–407. <https://doi.org/10.1146/annurev.genet.40.110405.090423>.
- Seshasayee ASN, Fraser GM, Luscombe NM. Comparative genomics of cyclic-di-GMP signalling in bacteria: post-translational regulation and catalytic activity. *Nucleic Acids Res*. 2010;38:5970–81. <https://doi.org/10.1093/nar/gkq382>.
- Randall TE, Eckart K, Kakumanu S, Price-Whelan A, Dietrich LEP, Harrison JJ. Sensory perception in bacterial cyclic diguanylate signal transduction. *J Bacteriol*. 2022;204:e00433–521. <https://doi.org/10.1128/jb.00433-21>.
- Römling U, Gomelsky M, Galperin MY. C-di-GMP: the dawning of a novel bacterial signalling system. *Mol Microbiol*. 2005;57:629–39. <https://doi.org/10.1111/j.1365-2958.2005.04697.x>.
- Krol E, Schäper S, Becker A. Cyclic di-GMP signaling controlling the free-living lifestyle of alpha-proteobacterial rhizobia. *Biol Chem*. 2020;401:1335–48. <https://doi.org/10.1515/hsz-2020-0232>.
- Jenal U, Reinders A, Lori C. Cyclic di-GMP: second messenger extraordinaire. *Nat Rev Microbiol*. 2017;15:271–84. <https://doi.org/10.1038/nrmicro.2016.190>.
- Brilli M, Fondi M, Fani R, Mengoni A, Ferri L, Bazzicalupo M, et al. The diversity and evolution of cell cycle regulation in alpha-proteobacteria: a comparative genomic analysis. *BMC Syst Biol*. 2010;4:52.
- Biondi EG, Reisinger SJ, Skerker JM, Arif M, Perchuk BS, Ryan KR, et al. Regulation of the bacterial cell cycle by an integrated genetic circuit. *Nature*. 2006;444:899–904. <https://doi.org/10.1038/nature05321>.
- Greene SE, Brilli M, Biondi EG, Komeili A. Analysis of the CtrA pathway in *Magnetospirillum* reveals an ancestral role in motility in alphaproteobacteria. *J Bacteriol*. 2012;194:2973–86. <https://doi.org/10.1128/jb.00170-12>.
- Poncin K, Gillet S, De Bolle X. Learning from the master: targets and functions of the CtrA response regulator in *Brucella abortus* and other alpha-proteobacteria. *FEMS Microbiol Rev*. 2018;42:500–13. <https://doi.org/10.1093/femsre/fuy019>.
- Lori C, Ozaki S, Steiner S, Bohm R, Abel S, Dubey BN, et al. Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. *Nature*. 2015;523:236–9. <https://doi.org/10.1038/nature14473>.
- Pallegar P, Peña-Castillo L, Langille E, Gomelsky M, Lang AS. Cyclic di-GMP-mediated regulation of gene transfer and motility in *Rhodobacter capsulatus*. *J Bacteriol*. 2020;202:e00554–619. <https://doi.org/10.1128/JB.00554-19>.



27. Koppenhöfer S, Lang AS. Interactions among redox regulators and the CtrA phosphorelay in *Dinoroseobacter shibae* and *Rhodobacter capsulatus*. *Microorganisms*. 2020;8:562.
28. Mann TH, Seth Childers W, Blair JA, Eckart MR, Shapiro L. A cell cycle kinase with tandem sensory PAS domains integrates cell fate cues. *Nat Commun*. 2016;7:11454.
29. Farrera-Calderon RG, Pallegar P, Westbye AB, Wiesmann C, Lang AS, Beatty JT. The CckA-ChpT-CtrA phosphorelay controlling *Rhodobacter capsulatus* gene transfer agent production is bidirectional and regulated by cyclic di-GMP. *J Bacteriol*. 2021;203:e00525–620. <https://doi.org/10.1128/JB.00525-20>.
30. Mercer RG, Callister SJ, Lipton MS, Pasa-Tolic L, Strnad H, Paces V, et al. Loss of the response regulator CtrA causes pleiotropic effects on gene expression but does not affect growth phase regulation in *Rhodobacter capsulatus*. *J Bacteriol*. 2010;192:2701–10. <https://doi.org/10.1128/jb.00160-10>.
31. Gowrishankar J, Harinarayanan R. Why is transcription coupled to translation in bacteria? *Mol Microbiol*. 2004;54:598–603. <https://doi.org/10.1111/j.1365-2958.2004.04289.x>.
32. Slager J, Veening J-W. Hard-wired control of bacterial processes by chromosomal gene location. *Trends Microbiol*. 2016;24:788–800. <https://doi.org/10.1016/j.tim.2016.06.003>.
33. Narula J, Kuchina A, Lee DD, Fujita M, Süel GM, Igoshin OA. Chromosomal arrangement of phosphorelay genes couples sporulation and DNA replication. *Cell*. 2015;162:328–37. <https://doi.org/10.1016/j.cell.2015.06.012>.
34. Lazazzera BA, Hughes D. Location affects sporulation. *Nature*. 2015;525:42–3. <https://doi.org/10.1038/nature15207>.
35. Marinus MG, Løbner-Olesen A. DNA methylation. *EcoSal Plus*. 2014;6:10.1128/ecosalplus.ESP-0003-2013. <https://doi.org/10.1128/ecosalplus.ESP-0003-2013>.
36. Reisenauer A, Shapiro L. DNA methylation affects the cell cycle transcription of the CtrA global regulator in *Caulobacter*. *EMBO J*. 2002;21:4969–77. <https://doi.org/10.1093/emboj/cdf490>.
37. Tomasch J, Koppenhöfer S, Lang AS. Connection between chromosomal location and function of CtrA phosphorelay genes in alphaproteobacteria. *Front Microbiol*. 2021;12:662907.
38. Sondermann H, Shikuma NJ, Yildiz FH. You've come a long way: c-di-GMP signaling. *Curr Opin Microbiol*. 2012;15:140–6. <https://doi.org/10.1016/j.mib.2011.12.008>.
39. PFAM Database BT - Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006. p. 1392. [https://doi.org/10.1007/3-540-29623-9\\_8198](https://doi.org/10.1007/3-540-29623-9_8198).
40. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49:D480–9. <https://doi.org/10.1093/nar/gkaa1100>.
41. Luo H, Quan C-L, Peng C, Gao F. Recent development of Ori-Finder system and DoriC database for microbial replication origins. *Brief Bioinform*. 2019;20:1114–24. <https://doi.org/10.1093/bib/bbx174>.
42. Rocha EPC. The replication-related organization of bacterial genomes. *Microbiology*. 2004;150:1609–27.
43. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>.
44. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28:1647–9. <https://doi.org/10.1093/bioinformatics/bts199>.
45. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37:1530–4. <https://doi.org/10.1093/molbev/msaa015>.
46. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evol*. 1985;39:783–91.
47. Sankoff D, Abel Y, Hein J. A tree a window a hill: generalization of nearest-neighbor interchange in phylogenetic optimization. *J Classif*. 1994;11:209–32. <https://doi.org/10.1007/BF01195680>.
48. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49:W293–6. <https://doi.org/10.1093/nar/gkab301>.
49. Cooper VS, Vohr SH, Wrocklage SC, Hatcher PJ. Why genes evolve faster on secondary chromosomes in bacteria. *PLOS Comput Biol*. 2010;6:e1000732. <https://doi.org/10.1371/journal.pcbi.1000732>.
50. Nagatani A. Phytochrome: structural basis for its functions. *Curr Opin Plant Biol*. 2010;13:565–70. <https://doi.org/10.1016/j.pbi.2010.07.002>.
51. Galperin MY, Gaidenko TA, Mulikidjanian AY, Nakano M, Price CW. MHYT, a new integral membrane sensor domain. *FEMS Microbiol Lett*. 2001;205:17–23. <https://doi.org/10.1111/j.1574-6968.2001.tb10919.x>.
52. Anantharaman V, Aravind L. Cache - a signaling domain common to animal Ca<sup>2+</sup>-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends Biochem Sci*. 2000;25:535–7. [https://doi.org/10.1016/S0968-0004\(00\)01672-8](https://doi.org/10.1016/S0968-0004(00)01672-8).
53. Zhulin IB, Nikolskaya AN, Galperin MY. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in bacteria and archaea. *J Bacteriol*. 2003;185:285–94. <http://jfb.asm.org/cgi/content/abstract/185/1/285>.
54. Dunin-Horkawicz S, Lupas AN. Comprehensive analysis of HAMP domains: implications for transmembrane signal transduction. *J Mol Biol*. 2010;397:1156–74. <https://doi.org/10.1016/j.jmb.2010.02.031>.
55. Mascher T, Helmann JD, Uden G. Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev*. 2006;70:910–38. <https://doi.org/10.1128/MMBR.00020-06>.
56. Kato M, Mizuno T, Shimizu T, Hakoshima T. Insights into multistep phosphorelay from the crystal structure of the C-terminal HPT domain of ArcB. *Cell*. 1997;88:717–23. [https://doi.org/10.1016/S0092-8674\(00\)81914-5](https://doi.org/10.1016/S0092-8674(00)81914-5).
57. Dutta R, Inouye M. GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci*. 2000;25:24–8. [https://doi.org/10.1016/S0968-0004\(99\)01503-0](https://doi.org/10.1016/S0968-0004(99)01503-0).
58. Galperin MY, Chou S-H. Sequence conservation, domain architectures, and phylogenetic distribution of the HD-GYP type c-di-GMP phosphodiesterases. *J Bacteriol*. 2022;204:e00561–621. <https://doi.org/10.1128/jb.00561-21>.
59. Pallegar P, Canuti M, Langille E, Peña-Castillo L, Lang AS. A two-component system acquired by horizontal gene transfer modulates gene transfer and motility via cyclic dimeric GMP. *J Mol Biol*. 2020;432:4840–55. <https://doi.org/10.1016/j.jmb.2020.07.001>.
60. Patterson DC, Ruiz MP, Yoon H, Walker JA, Armache J-P, Yennawar NH, et al. Differential ligand-selective control of opposing enzymatic activities within a bifunctional c-di-GMP enzyme. *Proc Natl Acad Sci*. 2021;118:e2100657118. <https://doi.org/10.1073/pnas.2100657118>.
61. Soler-Bistué A, Aguilar-Pierlé S, Garcia-Garcera M, Val M-E, Sismeyro O, Varet H, et al. Macromolecular crowding links ribosomal protein gene dosage to growth rate in *Vibrio cholerae*. *BMC Biol*. 2020;18:43. <https://doi.org/10.1186/s12915-020-00777-5>.
62. Soler-Bistué A, Mondotte JA, Bland MJ, Val M-E, Saleh M-C, Mazel D. Genomic location of the major ribosomal protein gene locus determines *Vibrio cholerae* global growth and infectivity. *PLOS Genet*. 2015;11:e1005156. <https://doi.org/10.1371/journal.pgen.1005156>.
63. Helmstetter CE, Cooper S. DNA synthesis during the division cycle of rapidly growing *Escherichia coli* Br. *J Mol Biol*. 1968;31:507–18. [https://doi.org/10.1016/0022-2836\(68\)90424-5](https://doi.org/10.1016/0022-2836(68)90424-5).

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.