In silico comparative analysis of GGDEF and EAL domain signaling proteins from the Azospirillum genomes

Alberto Ramírez Mata†, César Millán Pacheco†, José F. Cruz Pérez†, Martha Minjárez Sáenz† and Beatriz E. Baca*†

Abstract

Background: The cyclic-di-GMP (c-di-GMP) second messenger exemplifies a signaling system that regulates many bacterial behaviors of key importance; among them, c-di-GMP controls the transition between motile and sessile life-styles in bacteria. Cellular c-di-GMP levels in bacteria are regulated by the opposite enzymatic activities of diguanylate cyclases and phosphodiesterases, which are proteins that have GGDEF and EAL domains, respectively. Azospirillum is a genus of plant-growth-promoting bacteria, and members of this genus have beneficial effects in many agronomically and ecologically essential plants. These bacteria also inhabit aquatic ecosystems, and have been isolated from humus-reducing habitats. Bioinformatic and structural approaches were used to identify genes predicted to encode GG[DE]EF, EAL and GG[DE]EAL domain proteins from nine genome sequences.

Results: The analyzed sequences revealed that the genomes of A. humicireducens SgZ-5T, A. lipoferum 4B, Azospirillum sp. BS10, A. thiophilum BV-5T, A. halopraeferens DSM3675, A. oryzae A2P, and A. brasilense Sp7, Sp245 and A39 encode for 29 to 41 of these predicted proteins. Notably, only 15 proteins were conserved in all nine genomes: eight GGDEF, three EAL and four GGDEF-EAL hybrid domain proteins, all of which corresponded to core genes in the genomes. The predicted proteins exhibited variable lengths, architectures and sensor domains. In addition, the predicted cellular localizations showed that some of the proteins to contain transmembrane domains, suggesting that these proteins are anchored to the membrane. Therefore, as reported in other soil bacteria, the Azospirillum genomes encode a large number of proteins that are likely involved in c-di-GMP metabolism. In addition, the data obtained here strongly suggest host specificity and environment specific adaptation.

Conclusions: Bacteria of the Azospirillum genus cope with diverse environmental conditions to survive in soil and aquatic habitats and, in certain cases, to colonize and benefit their host plant. Gaining information on the structures of proteins involved in c-di-GMP metabolism in Azospirillum appears to be an important step in determining the c-di-GMP signaling pathways, involved in the transition of a motile cell towards a biofilm life-style, as an example of microbial genome plasticity under diverse in situ environments.

Keywords: Azospirillum, Diguanylate cyclases, Phosphodiesterases, Second messenger cyclic-di-GMP, Biofilm
analyses of proteins that are potentially important in the interactions of these bacteria with the host plant and in their adaptability to either terrestrial or aquatic environments [6–14].

It is now well established that bacteria in natural environments persist by forming biofilms [15]. Bacteria are able to sense and respond to ecologically distinct abiotic and biotic conditions [15]. These systems are necessary of these bacteria to adapt to changing environmental conditions and to enable survival in highly competitive habitats, such as the plant rhizosphere, soil or aquatic environments. In addition, usually, bacteria must efficiently colonize the root surface and other diverse surfaces, to exert their beneficial effect, which implies that understanding the bacterial traits required for biofilm formation is crucial to understanding the mechanisms involved in colonization. In particular, motility, which involves the flagellar apparatus and the chemotactic response to root exudates, appears to be an important colonization attribute, as do the formation of cell aggregates and the production of capsular polysaccharides. Indeed, several chemotaxis and aerotaxis operons have been identified in A. brasilense [6, 16, 17], and different mutant strains are defective in biofilm formation and root surface colonization [18, 19].

Similarly, some of the genetic determinants involved in cell aggregation and flocculation in A. brasilense and A. lipoferum lead to the differentiation of cyst-like cells [20, 21] and appear to be important for root colonization [22, 23]. Indeed, cell morphology had an effect on the associated root colonization, and importantly, capsular polysaccharides production in the surrounding cells was observed [19, 21]. Moreover, recent data support the hypothesis that the second messenger c-di-GMP (bis-(3′-5′)-cyclic-dimeric-guanosine monophosphate) plays a role in controlling the chemotactic response and, hence, biofilm formation in Azospirillum [24–26].

In several bacteria, c-di-GMP plays an important role in regulating the transition of the cells between a motile state and a sessile biofilm state. In addition, c-di-GMP is relevant in other bacterial functions, such as motility, chemotaxis, capsular polysaccharide formation, and cellulose synthesis [27, 28]. The synthesis and degradation of c-di-GMP are coordinated by the opposing activities of diguanylate cyclases (DGCs), which contain the GGDEF domain, and phosphodiesterases (PDEs), which harbor EAL or HD-GYP domains [28]. The GGDEF, EAL and HD-GYP domains were the first c-di-GMP modules to be identified using bioinformatics analyses, and these domains are widely distributed in bacterial genomes [28–30]. Hybrid proteins that contain both GGDEF and EAL domains have also been identified. In addition, other domains are often present, including sensory and regulatory modules, such as Per/Arnt/Sim (PAS), GAF, HAMP, REC, and MHTY, that modulate their enzymatic activities in response to external stimuli [28, 31]. The identification and functional characterization of c-di-GMP-associated proteins have indicated that the downstream signaling mechanisms of the c-di-GMP pathway might be versatile [32].

Various in silico analyses of proteins have demonstrated that such studies can contribute to the identification to the function of a specific protein; these analyses have been performed by using a growing number of bioinformatics resources [33]. This work reports an analysis performed on the genomes of nine Azospirillum strains: A. brasilense Sp245, A. brasilense Sp7, A. brasilense Az39, A. lipoferum 4B, Azospirillum sp. B510, and the recently sequenced genomes of A. thiophilum BV-S5, A. halopraeferens DSM 3675, A. oryzae A2P, and A. luscireducens SgZ-S5. We focused on identifying the genes encoding proteins containing the GG[D/E]EF, EAL and GG[D/E]EF-EAL domains and on characterizing the associated sensing and signaling domains. This work will contribute to our understanding of this important family of proteins that regulate cellular levels of c-di-GMP in Azospirillum.

**Methods**

We constructed the repertoire of genes coding for GG[D/E]EF, EAL and GG[D/E]EF-EAL domain proteins by analyzing nine Azospirillum genomes. The accession numbers of the chromosome and the other replicons in the genome of these strains are listed in Table 1.

The genomes were analyzed by performing BLAST searches of the GenBank database of the National Center for Biotechnology Information, with the Rapid annotation using subsystems technology (RAST) server [34], PFAM [35], SMART [36], PROSITE [37], conserved domain database [38], and MiST2.2 [39], websites to search for protein sequences from the Sp245, Sp7, Az39, 4B, B510, BV-S5, DMS3675, A2P and SgZ-S5 genome sequences. The amino acids of the motifs present in the various domains were identified using conserved domain database [38]. Clustal Omega was used to generate multiple protein sequence alignments [40, 41]. The localization of the signaling motifs was identified using the transmembrane helices in proteins (TMHMM) server to predict transmembrane helices [42]. All programs were used by following the specified parameters for successful analysis. The platform I-Tasser web server was used for automated protein structure and function predictions [43, 44]. Comparisons of the protein sequence domains were performed using the well-characterized homologous protein PleD of Caulobacter crescentus as a reference for the amino acid motifs of the GGDEF protein domain [45] and RocR from Pseudomonas aeruginosa as a reference for the EAL containing domain [46]. The crystal structures of a DGC (WspR) from P. aeruginosa [47] and the EAL domain of
| Strain   | Chromosome Size (Mbs) | Accession numbers of the genomes of the Azospirillum sp., strains used in this study and the predicted modular signaling proteins involved in c-di-GMP metabolism |
|----------|-----------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Strain   | Chr 1                 | Chr 2                                                                                                                                | Chr 3                                                                                                                             | Chr 4                                                                                                                             |Chr 5                                                                                                                             |
| A. bra. Sp245 | NC_016617.1          | 3.01                                                                                                                               | 2,690                                                                                                                             | 10/3/6                                                                                                                             |                                                                                                                                  |
| A. bra. Sp250 | NC_016594.1          | 1.76                                                                                                                               | 1748                                                                                                                             | 7/0/2                                                                                                                             |                                                                                                                                  |
| A. bra. Sp260 | NC_016618.1          | 0.912                                                                                                                              | 884                                                                                                                               | 1/1/0                                                                                                                             |                                                                                                                                  |
| A. bra. Sp270 | NC_016595.1          | 0.778                                                                                                                              | 808                                                                                                                               | 1/0/70                                                                                                                            |                                                                                                                                  |
| A. bra. Sp280 | NC_016596.1          | 0.690                                                                                                                              | 671                                                                                                                               | 1/1/2                                                                                                                             |                                                                                                                                  |
| A. bra. Sp290 | NC_016597.1          | 0.191                                                                                                                              | 162                                                                                                                               | -                                                                                                                                  |                                                                                                                                  |
| A. bra. Sp300 | NC_016619             | 0.612                                                                                                                              | -                                                                                                                                  | -                                                                                                                                  |                                                                                                                                  |
| A. bra. Sp310 | NC_016620             | 0.167                                                                                                                               | -                                                                                                                                  | -                                                                                                                                  |                                                                                                                                  |
| A. bra. Az39 | NZ_CP007793.1         | 3.06                                                                                                                               | 2767                                                                                                                              | 10/3/6                                                                                                                             |                                                                                                                                  |
| A. bra. Az40 | NZ_CP007794.1         | 1.9                                                                                                                                | 1610                                                                                                                             | 7/1/2                                                                                                                             |                                                                                                                                  |
| A. bra. Az41 | NZ_CP007795.1         | 0.819                                                                                                                              | 627                                                                                                                               | 1/0/0                                                                                                                             |                                                                                                                                  |
| A. bra. Az42 | NZ_CP007796.1         | 0.645                                                                                                                              | 555                                                                                                                               | 1/1/2                                                                                                                             |                                                                                                                                  |
| A. bra. Az43 | NZ_CP007797.1         | 0.206                                                                                                                              | 171                                                                                                                               | -                                                                                                                                  |                                                                                                                                  |
| A. bra. Az44 | NZ_CP007798.1         | 0.159                                                                                                                              | 110                                                                                                                               | -                                                                                                                                  |                                                                                                                                  |
| A. lip. 4B | NC_016622.1           | 2.99                                                                                                                               | 2693                                                                                                                             | 11/4/10                                                                                                                            |                                                                                                                                  |
| A. lip. 4C | NC_016585.1           | 1.04                                                                                                                               | 867                                                                                                                               | 1/0/0                                                                                                                             |                                                                                                                                  |
| A. lip. 4D | NC_016586.1           | 0.750                                                                                                                              | 619                                                                                                                               | 2/0/2                                                                                                                             |                                                                                                                                  |
| A. lip. 4E | NC_016623.1           | 0.648                                                                                                                              | 515                                                                                                                               | 3/1/2                                                                                                                             |                                                                                                                                  |
| A. lip. 4F | NC_016587.1           | 0.645                                                                                                                              | 555                                                                                                                               | 1/0/1                                                                                                                             |                                                                                                                                  |
| A. lip. 4G | NC_016588.1           | 0.478                                                                                                                              | 384                                                                                                                               | 1/0/1                                                                                                                             |                                                                                                                                  |
| A. lip. 4H | NC_016589.1           | 0.295                                                                                                                              | 220                                                                                                                               | -                                                                                                                                  |                                                                                                                                  |
| Azo B510 | NC_013854.1           | 3.3                                                                                                                                | 2994                                                                                                                             | 13/4/9                                                                                                                             |                                                                                                                                  |
| Azo B511 | NC_013855.1           | 1.46                                                                                                                               | 1127                                                                                                                             | 2/0/4                                                                                                                             |                                                                                                                                  |
| Azo B512 | NC_013856.1           | 0.723                                                                                                                              | 613                                                                                                                               | 1/0/0                                                                                                                             |                                                                                                                                  |
| Azo B513 | NC_013857.1           | 0.681                                                                                                                              | 532                                                                                                                               | 3/1/2                                                                                                                             |                                                                                                                                  |
| Azo B514 | NC_013858.1           | 0.628                                                                                                                              | 536                                                                                                                               | 0/0/1                                                                                                                             |                                                                                                                                  |
| Azo B515 | NC_013859.1           | 0.537                                                                                                                              | 422                                                                                                                               | 0/0/1                                                                                                                             |                                                                                                                                  |
| Azo B516 | NC_013860.1           | 0.261                                                                                                                              | 181                                                                                                                               | -                                                                                                                                  |                                                                                                                                  |
| A. halopraeferens DSM3675 | NZ_AUCF00000000.1 | 5.26                                                                                                                              | 4680                                                                                                                             | –                                                                                                                                  |                                                                                                                                  |
| A. oryzae A2P | NZ_FXAK00000000.1 | 7.67                                                                                                                              | 6669                                                                                                                             | –                                                                                                                                  |                                                                                                                                  |
| A. humicireducens SgZ-5 | NZ_CP015285.1 | 3.18                                                                                                                              | 2846                                                                                                                             | –                                                                                                                                  |                                                                                                                                  |

**Notes:**
- **Chr:** Chromosome
- **p1:** Plasmid 1
- **p2:** Plasmid 2
- **p3:** Plasmid 3
- **p4:** Plasmid 4
- **p5:** Plasmid 5
- **p6:** Plasmid 6
- **CDS:** Number of protein-coding genes
- **DGC/PDE/Hybrid:** Number of GGDEF, EAL and hybrids proteins per replicon
- **R:** References

**Table 1.** Accession numbers of the genomes of the Azospirillum sp. strains used in this study and the predicted modular signaling proteins involved in c-di-GMP metabolism.
C. crescentus were used for the structural analyses [48] as suggested by the server and the I-Tasser parameters were run [49, 50]. In this study, models exhibited a higher C-scores (better model) were used, and models analyzed against the structure with higher resolution were used to create the corresponding model. The analysis was conducted and figures were made using the Chimera UCSF program [49] and VDM [50].

Results
Number and features of domains identified in the translated products from genes encoding predicted DGC and PDE proteins
A search for genes encoding enzymes involved in c-di-GMP metabolism was performed in the genomes of three strains of A. brasilense (Sp245, Sp7 and Az39) and in the genomes of A. lipoferum 4B, Azospirillum spp. BS10, A. thiophilum BV-S-T, A. halopraeferens DM3675, A. oryzae A2P, and A. humicireducens Sg-Z-5T (Table 1). Some of these strains have composite genomes, containing several large plasmid-type replicons designated chromids [51] (because they contain essential genes), in addition to the chromosome (the largest replicon). A systematic analysis and comparison of the 9 genomes (Table 1) was performed as described in the methods section to identify the putative translated products that have diguanylate cyclase (DGC) and phosphodiesterase (PDE) activities and to define the amino acid motifs or signatures involved in catalytic activity, allosteric inhibition and interaction with metals (magnesium or manganese). This survey led to identification of three enzymatic classes of predicted proteins: DGCs, PDEs and hybrid DGC-PDEs. Indeed, even though the GGDEF and EAL domain-containing proteins have opposing activities, these two domains are often found coupled in the same proteins, which are referred to as hybrid proteins because they carry both domains. This nomenclature agrees with the presence of conserved amino acid motifs and with the distinctive secondary structure topology of these proteins [52]. The survey results led to the construction of a catalog of proteins predicted to be involved in c-di-GMP metabolism, as shown in Table 1. The number of genes encoding GGDEF, EAL, and hybrid domain proteins in each genome are as follows: A. humicireducens, 29 genes; A. brasilense Sp7, 34 genes; A. brasilense Sp245, Az39, and A. thiophilum, 35 genes each; A. halopraeferens 38 genes, A. lipoferum 4B and A. thiophilum 40 genes each; and Azospirillum spp. BS10 41 genes (Tables 1 and 2).

The relative proportions of enzymes from each of the three classes, is shown in Table 2. Azospirillum genomes encode for 14 to 20 DGCs enzymes, which contain the GGDEF domain (Table 2 and Additional file 1: Table S1). Four PDE enzymes, which contain the EAL domains, are present in A. brasilense Sp7, A. halopraeferens, A. oryzae, and A. humicireducens, whereas five PDE enzymes are found in the other strains (Table 2 and Additional files 1, 2 and 3: Tables S1, S2, S3, and S4). From the third class of enzymes, representing the hybrid proteins, 10 of these hybrid proteins are present in A. brasilense Sp245 and Az39 and nine in Sp7. The number of hybrid domain proteins is larger in the other six genomes, which contain 11 to 17 (Table 2 and Additional files 1 and 2: Tables: S2, S3, and S4). There is a core set of fifteen genes that are completely conserved among all nine Azospirillum strains; these genes encode for the following proteins: (i) eight DGCs, which belong to the PleD, WspR, or CdgA families [26] and seem to be functionally important independent of host-related or environmental specializations, in the Azospirillum strains, (ii) three PDEs, which include ChsA, a protein that was functionally characterized as being involved in chemotaxis and aerotaxis [25, 53]; and (iii) four highly conserved GGDEF-EAL hybrid proteins (Table 3 and Additional file 1: Table S1). The proteins of each class differ in their lengths, architectures, sensor domains and cellular localization, which were predicted by structural analysis (Table 3 and Additional file 3: Table S3). Almost all of the proteins harbor accessory domains at their N-termini, with a few exceptions that will be described later in the manuscript. A similar diversity in the number of proteins and architecture has also been shown in other bacterial genomes [54, 55]. In particular, as observed in the case of Azospirillum, proteins encoded by the genomes as containing EAL domains were less frequent than proteins encoded as containing GGDEF domains [54–56].

The subgroup of hybrid proteins containing tandem GGDEF/EAL or EAL/GGDEF domains are classified as DGCs, PDEs or DGC/PDEs (hybrid proteins) depending on the degree of conservation of the critical amino acid residues in their signature domains [28, 45, 46]. It was observed that all Azospirillum genomes have predominantly two types of hybrid proteins (Table 3 and Additional files 3 and 4: Tables: S3 and S5). In the first type of hybrid protein, the domains comprise highly conserved amino acid sequences suggesting that these hybrid proteins may exhibit both DGC and PDE activities. Because hybrid proteins also contain a regulatory partner or a sensor domain, these proteins require a mechanism by which to modulate their opposing activities. The associated sensor domain might determine the balance between the dual enzymatic activities via internal or extracellular signaling, as previously described for DcpA, a DGC/PDE protein from Agrobacterium tumefaciens that regulates attachment and biofilm formation. BphGL is a photoreceptor from Rhodobacter sphaeroides that is capable of both c-di-GMP synthesis and hydrolysis, and MucR is a protein from P. aeruginosa that has dual
activity and is involved in alginate biosynthesis via c-di-GMP signaling [57–59].

In contrast, there were hybrid proteins in each *Azospirillum* genome that belonged to the second type of hybrid protein and were predicted to be enzymatically inactive with a “highly degenerate” GGDEF domain (Additional file 2: Table S2). However, the EAL domains of these proteins contain all of the amino acid motifs for PDE activity and are highly conserved [28, 45, 46], which suggests that these proteins correspond to PDEs that are catalytically active. This second type of hybrid proteins, with only a conserved EAL catalytic site, usually also has a signal-sensing partner domain, suggesting distinct modes for the regulation of PDE activity under different contexts, as shown in Table 3 (Additional files 2, 3 and 4: Tables: S2, S3 and S5) and as previously reported [60, 61].

### Genomic relatedness between *Azospirillum* strains

Next, we performed a Venn diagram analysis in which each circle contained the memberships of the compared genomes. The relationships between all the putative proteins involved in c-di-GMP metabolism were assessed by considering whether they were conserved or were restricted to only one genome. The *A. brasilense* Sp245, Sp7 and Az39 genomes were compared with the genome of *A. lipoferum* 4B; 21 proteins were shared, but 18 proteins were predicted to be unique to the 4B genome (Fig. 1a). The B510 genome was compared with the genomes of the *A. humicireducens*, *A. thiophilum* and *A. oryzae* strains, and 23 proteins were conserved in all four genomes (Fig. 1b); only seven proteins were unique to the B510 genome, indicating that these genomes were the most closely related. Only 18 proteins were conserved in among the genomes of *A. brasilense* Sp245, Sp7, *A. lipoferum* 4B, and *A. halopraeferens*, and 13 proteins were found to be unique to the genome of *A. halopraeferens* (Fig. 1c). As these proteins are involved in signaling, these findings suggest that *Azospirillum* have evolved diverse transduction pathways, allowing better adaptation to a given niche.

### Sensory and regulatory module domains identified in GGDEF, EAL, and hybrid proteins

We further described the domain architectures found in each of the aforementioned sensory signaling domains identified in the GGDEF, EAL, and hybrid proteins; most of the proteins contained at least one predicted sensory domain, as shown in Table 3 (Additional files 2, 3 and 4: Tables: S2, S3 and S5). However, we found some exceptions, i.e., one to three putative GGDEF proteins, two EAL proteins, and one to two predicted GGDEF-EAL proteins were identified in all of the analyzed genomes. Other predicted proteins were seen to have at least one sensory domain. These regulatory or sensory domains detect small molecules, such as redox potential molecules, oxygen, nitric oxide (NO), light, voltage, osmolarity, and nutrients, and are also involved in protein-protein interactions. These domains enable the bacterium to integrate various types of input signals to establish a coordinated cellular output. In addition, all of these domains have regulatory functions that modulate the enzymatic activities of DGCs and PDEs in response to diverse environmental stimuli [28–32, 54, 55, 62, 63]. Indeed, the REC domain, which is a regulatory domain belonging to the Che-Y-like superfamily, has been identified as a receiver (phosphor-acceptor) domain or module that regulates the output of DGCs that responds to extracellular or intracellular signals transduced by their cognate sensor histidine kinases. The REC domain can be used to determine activity because its drives GGDEF dimerization, a process that is essential for activity and for cellular localization of the protein [28, 45, 64–67]. In addition, a careful examination of at least three databases revealed that some proteins were predicted to have transmembrane domains (TMD) and a peptide signal (signal P). These protein topologies suggested that the putative proteins are anchored to a membrane (Table 3 and Additional files 3 and 4: Tables: S3 and S5).

The PAS domain appears to be prevalent in proteins involved in c-di-GMP synthesis and degradation. In effect, as many as 38 predicted proteins contain one PAS/PAC domain or two or three domains in tandem, as is shown in Tables 3, 3S, and 5S. This domain is the most abundant sensory module found in signal transduction proteins throughout the bacterial kingdom; this domain generally binds small molecules and is the largest superfamily among domains solely dedicated to signal transduction [68]. Moreover, PAS/PAC domains are also involved in the protein-protein interactions that lead to

---

**Table 2** Frequency of occurrence of genes encoding GGDEF, EAL, and hybrid domain proteins in the genomes of *Azospirillum* strains

| DOMAIN | A. brasilense Sp245 | A. brasilense Sp7 | A. brasilense Az39 | A. lipoferum 4B | A. halopraeferens B510 | A. thiophilum BV-57 | A. halopraeferens DM53675 | A. oryzae A2P | A. humicireducens Sg-Z-5 |
|--------|------------------|-----------------|------------------|----------------|----------------------|------------------|---------------------|---------------|----------------------|
| GGDEF  | 20               | 20              | 20               | 19             | 19                   | 17               | 20                  | 19            | 14                   |
| EAL    | 5                | 5               | 5                | 5              | 4                    | 4                | 4                   | 4             | 4                    |
| Hybrid | 10               | 9               | 10               | 16             | 17                   | 13               | 14                  | 17            | 11                   |
| Total  | 35               | 34              | 35               | 40             | 41                   | 34               | 38                  | 40            | 29                   |

Data extracted from [http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins](http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins), SMART data base [http://smart.embl-heidelberg.de/](http://smart.embl-heidelberg.de/)
| DOMAIN ARCHITECTURE | A. brasilense SpP7 | A. brasilense Sp6 | A. brasilense Sp6A1 | Azospirillum 0100 | A. brasilense Sp83 | A. brasilense Sp72 |
|---------------------|--------------------|------------------|--------------------|-------------------|------------------|------------------|
| GGDEF-ONLY          |                    |                  |                    |                   |                  |                  |
| 1. Weil Family      |                    | *                | *                  |                   | *                |                  |
| 2.                  |                    | *                | *                  |                   |                  |                  |
| 3.                  |                    | *                | *                  |                   |                  |                  |
| 4.                  |                    | *                | *                  |                   |                  |                  |
| 5.                  |                    | *                | *                  |                   |                  |                  |
| 6. Weil Family      |                    | *                | *                  |                   |                  |                  |
| 7. CdgA Family      |                    | *                | *                  |                   |                  |                  |
| 8. Psh Family       |                    | *                | *                  |                   |                  |                  |
| 9.                  |                    | *                | *                  |                   |                  |                  |
| EAL-ONLY            |                    |                  |                    |                   | *                |                  |
| 10.                 |                    | *                | *                  |                   |                  |                  |
| 11.                 |                    | *                | *                  |                   |                  |                  |
| 12.                 |                    | *                | *                  |                   |                  |                  |
| GGDEF + EAL HYBRID  |                    |                  |                    |                   | *                |                  |
| 13.                 |                    | *                | *                  |                   |                  |                  |
| 14.                 |                    | *                | *                  |                   |                  | *                |
| 15.                 |                    | *                | *                  |                   |                  |                  |
| 16.                 |                    | *                | *                  |                   |                  | *                |
| 17.                 |                    | *                | *                  |                   |                  |                  |

Conserved predicted proteins containing GGDEF, EAL and GGDEF-EAL domains from all analyzed the genomes. Schematic representation of domain organization of proteins found in the genomes. The domain prediction was performed based on protein sequences derived from the genome sequences of CdgA (diguanylate cyclase A) and ChsA (phosphodiesterase) from the A. brasilense SpP7 strain using the simple modular architecture research tool (SMART) program. The GGDEF domains are shown in rose, and the EAL domains are shown in blue. The sensor domains predicted by SMART are shown as follows: PAS/PAC, represented the PAS-fold family (green); transmembrane domains, TMD (gray); CACHE 2, calcium channels and chemotaxis receptor family (blue); REC, response regulator receiver (blue). Data were extracted from http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins.

The website is indicated for each protein with its accession number according to BLASTP shown in Additional file 1: Table S1.
dimerization, which is usually essential for DGC activity [26, 28, 54, 55, 64–66]. In addition, PDE enzymes also often form dimers and tetramers, which are required for activity [28, 60, 69]. Thus, if the *Azospirillum* genes encoding DGCs or PDEs are expressed, they may be subjected to different environmental or intracellular signals associated with PAS/PAC domains.

Bioinformatic analysis also identified HAMP domains in some of the DGCs and DGC/PDEs in all *Azospirillum* species (Table 3, Additional files 3 and 4: Tables S3 and S5). These domains are defined as connectors between the periplasmic and cytoplasmic spaces. In addition, these domains transmit environmental stimuli across cytoplasmic membranes, and the conversion of that information to a signal triggers a change in function [70]. HAMP domains can be found in DGCs and DGC-PDEs that also have a TMD domain, suggesting that they are anchored to a membrane (Additional files 3 and 4: Tables S3 and S5).

**Analysis of features found in selected predicted proteins that are potentially involved in c-di-GMP signaling**

In our previous work, we identified the *chsA* gene, which encodes a PDE protein named ChsA that is involved in aerotaxis and chemotaxis [25, 53]. Herein, we found that the *chsA* gene is highly conserved in the genomes of all *Azospirillum* species (Table 3, Additional file 2: Table S2), showing a 99% to 51% similarity in amino acid residues in relation to *A. brasilense* Sp245. In addition, the DGC named CdgA, which is encoded by the ID: WP_035674663 gene in *A. brasilense* Sp7, was demonstrated to be involved in biofilm formation [26]. In this study, we found that *cgdA* is well conserved in the analyzed genomes (Table 3) and shares considerable identity (from 98 to 60%) with proteins from *A. brasilense* Sp245.

An interesting feature that was observed in the predicted DGCs and hybrid proteins from all nine genome sequences of the *Azospirillum* species is that they include both CACHE and TMDs in tandem domains (Table 3, and Additional files 3 and 4: Tables: S3 and S4). The CACHE (calcium channels and chemotaxis receptors) domain is an extracellular sensor domain that is present in bacteria and detects extracellular signals, such as small molecules and nutrients, and this domain is a ligand-binding domain commonly found in bacterial chemoreceptors [71]. These domains have been identified exclusively in proteins that contain output signaling domains, such as the DGC and PDE signal transduction proteins [71]. The majority of known ligands for the dCACHE_1 domain are amino acid sensors [72], whereas many of the single CACHE domains bind organic acids [73].

Only one highly conserved putative EAL-GGDEF hybrid protein was found in the studied genomes, and this protein had an EAL domain at the N-terminus (Additional files 3 and 4: Tables: S3 and S5). Both the EAL and the GGDEF domains are highly conserved, suggesting that these proteins might exhibit both activities depending on internal cellular signaling.

Several predicted DGCs and hybrid proteins contained both CHASE and TMDs domains at the N-termini (Additional files 3 and 4: Tables S3: and S5). The CHASE domain is a sensory domain named so because it is found in cyclases/histidine kinases with sensory functions. The CHASE domain is predicted to be a periplasmic domain consisting of 362 amino acid residues that serves as a transmembrane receptor and is often found in bacteria and plants. CHASE domains bind small molecules such as peptides and the phytohormone cytokinin [74]. As a bacterium associated with plants, *Azospirillum* might use this protein as a chemotaxis receptor.

---

**Fig. 1** Venn diagrams showing the distribution of families predicted to be involved in c-di-GMP metabolism from the genomes of *Azospirillum* strains. a *A. brasilense* Sp245, Sp7, and Az39 and *A. lipoforum* 4B. b *Azospirillum* sp. BS10, *A. thiophilum* BVS-5, *A. oryzae* A2P, and *A. humicireducens* SgZ-5T. c *A. brasilense* Sp245, Sp7, *A. lipoforum* 4B, and *A. halopraeferens* DSM3675. Numbers in black indicate the number of protein families; numbers in parentheses refer to the number of unique proteins in each genome that not found in the other genomes.
The predicted hybrid protein (GenBank: Sp245, WP_014199675; Sp7, WP_059399655, and Az39, WP_040138308) found in the *A. brasilense* genomes but not in other genomes showed an interesting structural architecture (Additional file 3: Table S3). This protein possesses an MHYT domain encompassing seven TMDs that are all localized at the N-terminus and has an MHYT motif consisting of four conserved amino acid residues (methionine, histidine, tyrosine, and threonine) that are predicted to be located near the outer face of the inner membrane. It has been suggested that the MHYT domain serves as a sensing domain [31]. The membrane topology of the MHYT domain indicates that the conserved residues of this domain can coordinate one or two copper ions, suggesting that this domain plays a role in sensing oxygen, CO or NO. In addition, the C-terminus includes PAS-GGDEF-EAL domains. The gene encoding this protein is often fused to a LysR-type DNA-binding helix-turn-helix protein, and an investigation of the genome organization showed that the genomic position of this protein was conserved in the genomes of all the *A. brasilense* strains [31, 39]. In *P. aeruginosa*, alginate biosynthesis, formation of highly structured biofilms, and inhibition of swarming motility are regulated by MucR, which is a hybrid MHYT-DGC-PDE protein [75].

### Structural and tridimensional topographies

**Structural features of the GGDEF and EAL domains of the DGCs, PDEs and hybrids proteins from the *A. brasilense* Sp7 genome**

Next, we determined that the structural features of the predicted proteins found in the *A. brasilense* Sp7 genome confirm previous predictions. All models of the GGDEF, EAL or GGDEF/EAL proteins had C-scores higher than 0.49 (better models had values close to 2) [43, 44]. The C-values of all models are reported in Table 4.

**GGDEF-only DGCs and hybrid (GGDEF domain) proteins**

The GGDEF protein models were constructed based on the crystal structure of the GGDEF conserved domain of WspR (PDB id: 3BRE), which has a crystallographic resolution of 2.4 Å [47]. Structural alignments, shown in Fig. 2a, were performed using the GGDEF domain of 3BRE from amino acids L170 to Q339. Even though the putative DGC protein (GG[DE]EF-only domain) sequences analyzed did not possess a highly identity percentages (ranging from 30.41 to 48.02%) (Additional file 5: Table: S7a), the proteins contained all the essential conserved amino acid residues that bind the substrate, GTP, to have enzymatic activity [76, 77]. The Web Logo [78] alignment in Fig. 2a shows the characteristic

| GGDEF only | C score | EAL only | C score | GGDEF hybrid proteins | C score | EAL hybrid proteins | C score |
|------------|---------|----------|---------|-----------------------|---------|---------------------|---------|
| WP_059399097 | 0.79 | WP_059398606 | 1.53 | WP_079285130 | 1.25 | WP_059399067 | 0.61 |
| WP_051140034 | 1.53 | WP_051140161 | 1.61 | WP_051140186 | 1.33 | WP_079285130 | 1.83 |
| WP_035675850 | 0.58 | WP_035672792 | 1.08 | WP_051140104 | 1.28 | WP_051140186 | 1.80 |
| WP_035672942 | 0.60 | WP_035682417 | 1.83 | WP_059398931 | 1.18 | WP_051140104 | 1.78 |
| WP_035674663 | 0.58 | CAJ18244 | 1.60 | WP_059399067 | 0.68 | WP_059398931 | 1.73 |
| WP_035674304 | 0.59 | | | WP_051140268 | 1.24 | WP_051140528 | 1.82 |
| WP_035671267 | 0.59 | | | WP_035678503 | 1.27 | WP_035678503 | 1.65 |
| WP_035671094 | 0.49 | | | WP_059399655 | 1.27 | WP_059399655 | 1.72 |
| WP_035671042 | 1.08 | | | WP_059399677 | 1.29 | WP_059399677 | 1.71 |
| WP_035670844 | 0.59 | | | | | | |
| WP_035670654 | 0.79 | | | | | | |
| WP_035676633 | 0.61 | | | | | | |
| WP_035671246 | 0.73 | | | | | | |
| WP_051140383 | 0.81 | | | | | | |
| WP_035678542 | 0.88 | | | | | | |
| WP_059399331 | 0.52 | | | | | | |
| WP_059399449 | 1.15 | | | | | | |
| WP_035682812 | 0.96 | | | | | | |
| WP_079285367 | 0.54 | | | | | | |
| WP_051140397 | 0.77 | | | | | | |

The data were extracted from the I-TASSER server for protein structure and function. The website is indicated for each gene with is accession number according to BLASTP.
secondary structure elements of the DGCs, such as five α helices and seven short β strands (α1β1α2α3β2α4β3-β6α5β7) \[52\]. In addition, regions of the sequence were identified with high root mean square deviation (RMSD) values, namely, R195 to L203, C240 to L246, P260 to P264, and T275 to F312, as shown in Additional file 5: Table: S7a. High RMSD values were found for most of these regions except for T275 to F312, which corresponded to the loop regions shown in Fig. 2a. The region from T275 to F312 corresponded to a loop region and a β strand that crossed from one side to the other in each of the protein models. In this region, the models exhibited a secondary structure. We suggested that the absence of a secondary structure may be the result of protein model construction and that these structures must be studied by circular dichroism spectroscopy or dynamic simulations in further studies.

The GGDEF hybrid proteins had a sequence conservation percentages ranging from 24.11 to 33.14% (Additional file 5: Table: S7b). As indicated by the high RMSD values, these regions corresponded to sequence gaps or insertions, primarily from R195 to Q202 and S280 to L294. The ALJ36098 protein (GenBank ID: WP_059398931) was found to have the most different
sequence in the characteristic of GGDEF motif (Additional file 2: Table: S2 b3). Indeed, ALJ36098 had an SDHAF motif; this variation is divergent in the sizes and charges of the amino acid residues involved in enzymatic activity, suggesting that this predicted protein lacks catalytic activity \[64–66, 76, 77\]. Additionally, an electrostatic potential analysis mapped on the protein surfaces showed that the SDHAF motif changed the charge distribution. These changes may confer on the protein a different affinity for its ligand compared to the affinity of GGDEF proteins (Fig. 3) \[76, 77\].

**EAL-only PDEs and hybrid (EAL domain) proteins**

**EAL-only proteins**

EAL proteins from the *A. brasilense* Sp7 genome were compared to the crystal structure of the EAL domain of PdeA from *C. crescentus* (PDB ID: 3U2E), which has a crystallographic resolution of 2.32 Å \[48\]. When compared to 3U2E, the EAL-only proteins had lower sequence conservation, ranging from 19.50 to 36.44%, than the EAL hybrid proteins. (Additional file 5: Table: S7c). As indicated in Fig. 1b, the regions with high RMSD-backbone values are located in different sections across the structures, e.g., the first ten amino acids at the amino terminus; these regions are A327 to G350, W369 to P393 and R464 to V501. As indicated in Fig. 1b, some loops from the EAL-only proteins with visible discrepancies were included in the comparison to the crystallographic structure. However, as shown in Fig. 2b, the crystal structures of the EAL domains show that the proteins possess the conserved signature motif (EAL) in addition to the flexible loop (“loop 6”), which has been extensively characterized in (β/α) barrel proteins, and that both of these features are required for catalytic activity in a functional protein \[60, 69, 76, 79, 80\].

**EAL-hybrid proteins**

The sequence conservation percentages for these proteins ranged from 29.02 to 43.20%. The lowest conservation percentage corresponded to ALJ36617 (GenBank ID: WP_059399067) and the highest to ALJ36098 (GenBank ID: WP_0593998931). Moreover, the low sequence conservation appeared not to be a factor for model prediction by I-Tasser. As shown in Fig. 2c, the structural alignment of 3U2E against all of the EAL hybrid proteins gave a low RMSD value for the backbone atoms. The amino acid residues, when present, had high RMSD values for residues from T368 to T380 (3U2E sequence numbering) (Additional file 5: Table: S7d). These regions presented greater sequence fluctuation, including gaps in some of the sequences; however, the amino acid residues involved in the ligand-metal binding of c-di-GMP and enzymatic activity are highly conserved in all sequences \[48, 60, 76, 79–81\] (Fig. 2c). In addition, residue M358 exhibited some changes, including change from a methionine in 3U2E to hydrophobic residues (such as valine, leucine or alanine) in a majority of the analyzed sequences and to threonine in ALJ39102 (GenBank ID: WP_059399677).

**Discussion**

It is well documented that in the bacterial kingdom c-di-GMP signaling is linked to biofilm formation and several other phenotypes that are important to the lifestyle of bacteria. We advanced our understanding of c-di-GMP signaling in the most important species of *Azospirillum*, e.g., those that are used as inoculants to promote plant growth or in soil bioremediation, by studying how many domain architectures and tridimensional structures were contained in the predicted proteins of genes encoding DCGs, PDEs and DGC-PDEs; these genes are widespread and are found in other environmental and soil bacteria, such as *Sinorhizobium meliloti* \[82\], other species of *Rhizobium* \[83\], *Pseudomonas putida* \[62\], *Shewanella*
oneidensis [63, 84], and Burkholderia lata SK875 [85], which reportedly encode a considerable number of proteins predicted to be involved in c-di-GMP metabolism.

The comparisons of the genomes showed that 15 proteins shared a significant percentage of identity at the amino acid residue level (the genes comprising the core), indicating the genetic relatedness among Azospirillum strains as previously described [6, 67, 86]. In addition, some of the genes were duplicated and identified at different genetic locations (chromosome or chromids) in the same genome, indicating that they might be derived from the duplication of a common ancestral gene that then diverged from the parent copy by mutation and selection, as proposed by the phylogenetic analysis. This evolution suggests that these genes were paralogs and that they were likely acquired by horizontal transfer (HGT), as defined for the A. brasilense and A. lipoferum genomes [6, 67, 86, 87]. Notably, it was observed these genomes possessed genes that encoded for ChsA, a PDE involved in chemotaxis and aerotaxis, and it has been well established by Russell et al. [25] that Azospirillum uses chemotaxis to navigate through the soil to find optimal surroundings for survival. Thus, the control of cellular motility by c-di-GMP signaling is the best illustration, to date, of the importance of a c-di-GMP-controlled rapid response to changing environmental conditions. Based on with phylogenetic analysis, the genomes were clustered in three groups: strains of A. brasilense (Sp245, Sp7 and Az39) included in the same clade; strain A. lipoferum 4B, which clustered with Azospirillum B510, A. humicireducens SgZ-5T, A. thiophilum BV-S5T, and A. oryzae A2P; and A. halopraeferens DSM3675, the genes of which were the most divergent. This is in agreement with previous studies on the whole-genomes of Azospirillum strains [6–14]. In addition, it was interesting to note that proteins encoded by the A. halopraeferens genome showed very complex structural features, as indicated in Table 5S (Additional file 4). A. halopraeferens, isolated from the rhizoplane of Kallar grass (Leptochloa fusca L. Kunth), is a salt tolerant bacterium [88]. The bacterium was inoculated to an oilseed halophyte Salicornia bigelovii Torr plant in salt-contaminated, infertile areas, and under these detrimental stress conditions, the plant-growth-promotion was significantly improved [89]. Therefore, the data obtained here strongly suggest host specificity and environment-specific adaptation.

The domain architectures of the deduced amino acid sequences of DGC and PDE proteins from Azospirillum genomes were also predicted that to included diverse sensor domains, such as REC, PAS/PAC, CHASE, GAF, MHYT and CACHE, that are involved in activity regulation by driving the protein dimerization process, which is essential for activity, or by sensing small molecules commonly found in rhizospheric or aquatic habitats [28–32, 69, 90]. These predictions are useful to predict how bacteria are able to monitor the internal metabolic status of a cell as well sense environmental cues, such as those from root exudates or the rhizosphere, or signals associated with a particular environment [57–59, 62, 63, 83, 91].

Furthermore, the cellular localization of some proteins was assessed by the presence or absence of transmembrane helices; the cellular localization of these proteins might indicate that the cellular c-di-GMP pool is localized to support functional micro-compartmentalization, which may be participate in the response to different environmental signals or may allow membrane localization of the protein after a spatial signal is sensed, thereby regulating its enzymatic activity with the corresponding co-localization of DGC, as previously described in several studies [28, 57, 59, 65, 77].

The analysis of structural architecture has proven to be very informative with regard to putative functions of signalling proteins. It was mentioned that a majority of hybrid proteins have conserved EAL catalytic sites, and these predicted proteins contribute to the total PDE cellular activity. The inactive GGDEF domains of these proteins function by regulating hydrolytic activity, as previously described in Xanthomonas oryzae pv. oryzae [91], or by acting as “trigger enzymes” with a dual function of either hydrolyzing c-di-GMP, or acting as an effector that binds to a transcriptional regulator that acts on a promotor involved in a signaling cascade. This signaling cascade controls matrix production in the biofilm; or can also control its own transcription, as described for PdeR and PdeL from Escherichia coli [92, 93]. This is the case for the ALJ36098 protein from A. brasilense Sp7 which is predicted to lack catalytic activity [76, 77]; however, this protein might exhibit regulatory or effector functions by binding c-di-GMP or GTP, as previously described for some hybrid proteins [70, 89–92], or may act as an effector in the c-di-GMP signaling cascade as previously described [92–94]. Considering the data presented here, we suggest that this structural analysis provides important information to predict the function of these proteins containing GGDEF, EAL, and hybrid domains, and creates a paradigm for future studies on the evolution of enzymes involved in c-di-GMP metabolism.

**Conclusions**

In summary, compared to other plant-associated bacteria Azospirillum were found to contain a number of similar genes (29 to 41) encoding DGCs and PDEs in their genomes. Our findings help elucidate the functions of the predicted hybrid multi-domain proteins, which allow the bacteria to integrate different signals via significant signaling plasticity. Indeed, this significant
flexibility might reflect differentially regulated c-di-GMP signaling mechanisms in *Azospirillum* that enable responses to distinct environmental and cellular signals. Therefore, in silico analysis of the ligand binding domains in the genomic sequences is a pre-requisite for further experimental characterization and evaluation of biological function. Thus, these conserved signaling proteins might be ecologically relevant and may explain how *Azospirillum* adapts to its specific ecological niche. An interesting question is raised regarding involvement of these proteins in physiological regulation. Future phenotypic and biochemical studies are needed to answer this question.

**Additional files**

**Additional file 1 Table S1.** Accession numbers of GGDEF, EAL and hybrid proteins encoded by genes conserved in all analyzed *Azospirillum* genomes. Data extracted from http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins. (DOCX 24 kb)

**Additional file 2 Table S2.** Accession numbers and classifications of GGDEF, EAL and hybrid proteins predicted based on the conservation of signature motifs that were found in all analyzed genomes. Data extracted from http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins and http://smart.embl-heidelberg.de/ following the notation from Römling et al. (2017). (DOC 109 kb)

**Additional file 3 Table S3 and Table S4.** Table S3. Repertoire of GGDEF, EAL and GGDEF-EAL predicted proteins, organization, and domain architectures found in the selected *Azospirillum* spp. genomes. Table S4. Accession numbers of GGDEF, EAL and hybrid proteins encoded by genes of select the analyzed *Azospirillum* genomes. Data extracted from http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins and http://smart.embl-heidelberg.de/ (DOCX 268 kb)

**Additional file 4 Table S5 and Table S6.** Table S5. Repertoire of GGDEF and GGDEF-EAL (hybrid) predicted proteins, organization, and domain architectures represented exclusively in the A. halopraeferens, A. thiophilum, and A. oryzae genomes. Table S6. Accession numbers of GGDEF and hybrid proteins encoded by genes exclusively found in A. halopraeferens, A. thiophilum, or A. oryzae genomes. Data extracted from http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins and http://smart.embl-heidelberg.de/ (DOCX 162 kb)

**Additional file 5 Table S7.** The alignments, root main square deviations (RMSDs) and sequence conservation percentages of proteins with GGDEF/EAL, and hybrid domains encoded by genes found in the *A. brasilense* Sp7 genome. Table 7Sa data including the GGDEF/EAL proteins; Table 7Sb, data including the EAL proteins; Table 7Sc and 7Sd, data including the hybrid proteins. Data extracted from http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins. (DOCX 6071 kb)

**Abbreviations**

BLAST: Basic local alignment search tool; c-di-GMP: Cyclic-di-GMP; DGCs: Diguanylate cyclases; MIST2: Microbial signal transduction database; PDEs: Phosphodiesterases; PFAM: Protein domain database; PGPB: Plant-growth-promoting bacteria; PROSITE: Protein domain database for functional characterization; RAST: Rapid annotation using subsystems technology; RNASD: Root main square deviations; SMART: Simple modular architecture research tool; VMD: Visual molecular dynamics

**Acknowledgements**

The authors are grateful to Dr. C. Elmerich, of Institute Pasteur Paris, France, for her constructive suggestions. We would like to thank all of the reviewers and the Senior Editor for their analysis of our manuscript and for their very helpful recommendations.

**Author contributions**

The work presented here was carried out as a collaboration among all authors. B.E.B conceived the study, supervised the research, and drafted the manuscript. A.R.M and, C.O.M supervised and participated in the genomic and structural analyses, and revised the manuscript. J.F.C.P and M.M.S retrieved the data from the database, performed the genomic and phylogenetic analyses and revised the manuscript. All authors read and approved the final manuscript.

**Funding**

This research project including the design of the study, data retrieval, analysis, data interpretation, and the writing of the manuscript, was funded by the Consejo Nacional de Ciencia y Tecnología (CONACyT), grant CB201-1549142, Programa para el mejoramiento del profesorado, grant PROMEP/103.5/13/8869, and the financial support of Vicenctorial de investigación y estudios de posgrado (VIPE). JFCP and MMS received a scholarship from CONACyT.

**Availability of data and materials**

All data generated or analyzed during this study are presented within the manuscript and/or additional files.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

All authors declare that they have no competing interests.

**Publisher’s Note**

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

**Author details**

1Centro de Investigaciones en Ciencias Microbiológicas, Benemérita Universidad Autónoma de Puebla. Edif. IC11, Ciudad Universitaria, Col. San Manuel Puebla Pue, CP72570 Puebla, Mexico. 2Facultad de Farmacia, Universidad Autónoma del Estado de Morelos, Av. Universidad #1001, Col. Chamilpa, CP, 62209 Cuernavaca, Morelos, Mexico.

Received: 4 September 2017 Accepted: 9 February 2018

**Published online:** 09 March 2018

**References**

1. Bashan Y. De-Bashan LE, Prabhu SR. Hernandez JP. Advances in plant growth-promoting bacterial inoculant technology: formulations and practical perspectives (1998–2013). Plant Soil. 2014;378:1–33.
2. Ökon Y, Labandera-Gonzalez CA. Agronomic applications of *Azospirillum* an evaluation of 20 years worldwide field inoculation. Soil Biol Biochem. 1994; 26:1591–601.
3. Lavrinenko K, Chemousova E, Grishina E, Dubinina G, Akimov V, Kuever J, et al. *Azospirillum* thiophilum sp. nov., a diazotrophic bacterium isolated from a sulfide spring. Int J Syst Evol Microbiol. 2010;60:2832–7.
4. Zhou S, Han L, Wang Y, Yang G, Li Zhuang L, Hu P. *Azospirillum humicireducens* sp. nov., a nitrogen-fixing bacterium isolated from a microbial fuel cell. Int J Syst Evol Microbiol. 2013;63:2618–24.
5. Tarrand J, Krieg N, Döbereiner J. A taxonomic study of the *Spirillum* lipoforum group, with descriptions of a new genus, *Azospirillum* gen. Nov., and two species, *Azospirillum* lipoforum (Beijerinck comb. nov. and *Azospirillum* brasilense sp. nov. Can J Microbiol. 1978;24:967–80.
6. Winniewski-Dyè F, Boratok K, Khalsa-Moyers G, Alexandre G, Sukharnikov LO, et al. *Azospirillum* genomes reveal transition of *Bacteria* from aquatic to terrestrial environments. PLoS Genet. 2011;7:e1002430. https://doi.org/10.1371/journal.pgen.1002430.
7. Shin YG, et al. School of Applied Biosciences, College of Agriculture and Life Sciences, Kyungpook National University. 2015. http://www.ncbi.nlm.nih.gov/muccore/
8. Rivera D, Revale S, Molina R, Gualpa J, Puente M. et al. Complete genome sequence of the model rhizosphere strain *Azospirillum brasilense* Ax29, successfully applied in agriculture. Genome Announc. 2014;2:e0083-14.
9. Kaneko T, Minamisawa K, Iwata T, Nakatsuka H, Mitsu H, et al. Complete genomic structure of the cultivated rice endophyte Azospirillum sp. BS10. DNA Res. 2010;17:37–50.

10. Fomenkov A, Vincze T, Grabovich M, Anton BP, Dubinina G, et al. Complete Genome Sequence of a Strain of Azospirillum thiophorum Isolated from a Sulfide Spring. Genome Announc. 4(1):e01522–15. https://doi.org/10.1128/genomeA.01523–15.

11. Kvak Y, Shin JH. First Azospirillum genome from aquatic environments: whole genome sequence of Azospirillum thiophorum BS-V5 a novel diazotroph harboring a capacity of sulfur chemolithotrophy from sulfide spring. Mar Genomics. 2016;25:21–4.

12. Kyrpides N, Huntemann M, Han J, Chen A, Mavromatis K, Markowitz V, et al. Azospirillum halopraeferens DSM 3675 G472 DRAFT_scaffold00001.C, whole genome shotgun sequence. https://www.ncbi.nlm.nih.gov/genome/annotation_prok/.

13. Zhou S, Han L, Wang Y, Yang G, Li Zhuang L, Azospirillum HP. humicrocides strain SgZ-3, complete. Genome. 2016; https://www.ncbi.nlm.nih.gov/genome/annotation_prok/.

14. Varghese N. Azospirillum oxyae strain AJP, whole genome shotgun sequence. 2017. https://www.ncbi.nlm.nih.gov/genome/annotation_prok/.

15. Jonas C, Meleiros O, Römling U. Regulation of c-di-GMP metabolism in biofilms. Future Microbiol. 2009;4:341–58. https://doi.org/10.2217/fmb.09.7.

16. Bible AN, Stephens BB, Ortega DR, Xie Z, Alexandre G. Analysis of chemo taxis-like signal transduction pathway in modulating motility, cell clumping, and cell length in the alphaproteobacterium Azospirillum brasilense. J Bacteriol. 1996;190:659–604.

17. Pereg Gerk L, Paquelin A, Gounon P, Kennedy IR, Elmerich C. A transcriptional regulator of the LuxR-UhpA family, Fca, controls flocculation and wheat root surface colonization by Azospirillum brasilense Sp7. Mol Plant-Microbe Interact. 1998;11:177–87.

18. Sadasivan L, Neera CA. Flocculation in Azospirillum brasilense and Azospirillum lipoflavum: exopolysaccharides and cyst formation. J Bacteriol. 1985;163:716–23.

19. Bashar Y, Mitiku G, Whitmoyer RE, Lepanehy H. Evidence that fibrillar anchoring is essential for Azospirillum brasilense cell attachment to sand. Plant Soil. 1991;132:73–83.

20. Burdman S, Jurkevitch E, Okon Y. Surface characteristics of Azospirillum brasilense in relation to cell aggregation and attachment to plant roots. Crit Rev Microbiol. 2000;26:291–110.

21. Fibach-Paldi S, Burdman S, Okon Y. Key physiological properties contributing to rhizosphere adaptation and plant growth promotion abilities of Azospirillum brasilense. FEMS Microbiol Lett. 2012;326:99–104.

22. Drogue B, Sanguin H, Bofarlan S, Pitgent-Combaret C, Wisniewski-Dyè F. Genome wide profiling of Azospirillum lipoflavum 4B gene expression during interaction with rice roots. doi: https://doi.org/10.1111/j.1574-6941.2014.02244.

23. Russell MH, Bible AN, Fang X, Gooding JR, Campagna SR, Gomelsky M, Alexandre G. Integration of the second messenger c-di-GMP into the chemotactic signaling pathway. MBio. 2013;4:e00001-13.

24. Ramírez-Mata A, López Lara LI, Quiroz-Vázquez ML, Romero Osorio A, Saúl Jijón-Moreno S, Baca BE. The cyclic-di-GMP diguanylate cyclase CdgA has a role in biofilm formation and exopolysaccharide production in Azospirillum brasilense Research Microbiol. 2016; https://doi.org/10.1016/j.resmic.2015.12.004.

25. Alexandre G. Chemotaxis control of transient cell aggregation. J Bacteriol. 2015;197:3230–7. https://doi.org/10.1128/9780.0121-15.

26. Römling U, Galperin MY, Gomelsky M. Cyclic-di-GMP: the first 25 years of a universal bacterial second messenger. Microbiol Mol Biol Rev. 2013;77:71–52.

27. Schirmer T, Jenal U. Structural and mechanistic determinants of c-di-GMP signalling. Nat Rev Microbiol. 2009;7:24–35.

28. Römling U, Galperin MY, Gomelsky M. Cyclic-di-GMP: the dawning of a novel bacterial signalling system. Mol Microbiol. 2005;57:629–39.

29. Galperin MY, Nikolayeva AN, Koonin EV. Novel domains of the prokaryotic two-component signal transduction systems. FEMS Microbiol Lett. 2001;203:11–21.

30. Mills E, Pulz IS, Kulaekara HD, Miller SI. The bacterial second messenger c-di-GMP: mechanisms of signalling. Cell Microbiol. 2011;13:1122–9.
58. Tarutina M, Ryenkov DA, Gomesky D. An unorthodox bacteriophytochrome from Rhodobacter sphaeroides is involved in turnover of the second messenger c-di-GMP. J Biol Chem. 2006;281:134751-8.

59. Wang Y, Hay ID, Rehrman ZU, Rehm BHA. Membrane-anchored MuCR mediates nitrate-dependent regulation of alginic production in Pseudomonas aeruginosa. Appl Microbiol Biotechnol. 2015;99:7253–65.

60. Chen MW, Kotaka M, Vonhejen C, Borregaard N, Rasmussen S, et al. Structural insights into the regulatory mechanism of the response regulator RocR from Pseudomonas aeruginosa in cyclic di-GMP signalling. J Bacteriol. 2012;194:4837–46.

61. Christen M, Christen B, Folcher M, Schaar T, Jentsch T, et al. Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GDP. J Biol Chem. 2002;277:30820–37.

62. Österberg S, Åberg A, Herrera-Seitz MK, Wolf-Watz M, Shingle V. Genetic dissection of a motility-associated c-di-GMP signalling protein of Pseudomonas putida. Environ Microbiol. 2013;15:556–66.

63. Sundaraman A, Kurosky J, Yan T, Klingenman GM, Joachimiak MP, et al. Shenawanella oneidensis MR-1 sensory box protein involved in aerobic and anaerobic growth. Appl Environ Microbiol. 2011;77:4647–56.

64. De N, Navarro MV, Navarro MV, Raghavan RV, Sondermann H. Determinants for the formation and dimerization of the phosphodiesterase active site of the cyclic di-GMP phosphodiesterase of Shewanella putrefaciens. J Biol Chem. 2007;282:29170–7.

65. Christen B, Folcher M, Schaar T, Jentsch T, Jentsch T, et al. Identification of a chemoreceptor for C2 and C3 carboxylic acids. Appl Environ Microbiol. 2015;81:5449–57.

66. Henry JT, Crosson S. Ligand-binding PAS domains in a genomic, cellular, and structural context. Annu Rev Microbiol. 2011;65:261–86. https://doi.org/10.1146/annurev-micro-121809-151631.

67. Phippen CW, Mikolajek H, Schlaefli HG, Keevil CW, Webb JS, Tews I. Structural insights into proteins. J Bacteriol. 2016;198:4631–46. https://doi.org/10.1128/JB.00790-16.

68. Paul R, Abel S, Wassmann P, Beck A, Hengge R. The EAL domain phosphodiesterase PdeB, a cyclic di-GMP-specific phosphodiesterase gene in Burkholderia lato SK875 attenuates virulence and enhances biofilm formation. J. Bacteriol. 2013;195:3827–33.

69. Wang Y, Xu J, Amin J, Wang Y, Zhu J, et al. Cyclic-di-GMP and EAL proteins play different roles in the control of Sinorhizobium meliloti growth, motility, exopolysaccharide production, and competitive nodulation on host alfalfa. Acta Biochim Biophys Sin. 2010;42:410–41.

70. Gao S, Romdhane SB, Beulens S, Kauter J, Lambrecht L, et al. Genomic analysis of cyclic-di-GMP-related genes in rhizobial type strains and functional analysis in Rhizobium etli. Appl Microbiol Biotechnol. 2014;98:6489–602.

71. Zhang H, Zhang J, Jia Y, Zhang W, Liu W, et al. CHASE domain. FEBS Lett. 2004;576:287–92.

72. Österberg S, Åberg A, Herrera-Seitz MK, Wolf-Watz M, Shingle V. Genetic dissection of a motility-associated c-di-GMP signalling protein of Pseudomonas putida. Environ Microbiol. 2013;15:556–66.

73. Sundaraman A, Kurosky J, Yan T, Klingenman GM, Joachimiak MP, et al. Shenawanella oneidensis MR-1 sensory box protein involved in aerobic and anaerobic growth. Appl Environ Microbiol. 2011;77:4647–56.

74. De N, Navarro MV, Baghavan ROW, Sondermann H. Determinants for the activation and autoinhibition of the diguanylate cyclase response regulator WspR. J Mol Biol. 2003;339:619–33.

75. Paul R, Seiler S, Arnott NC, Chan C, Schaar T, Jentsch T, et al. Identification of a chemoreceptor for C2 and C3 carboxylic acids. Appl Environ Microbiol. 2015;81:5449–57.

76. Christen M, Christen B, Folcher M, Schaar T, Jentsch T, et al. Identification of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GDP. J Biol Chem. 2002;277:30820–37.

77. Paul RW, Chen MW, Kotaka M, Vonhejen C, Borregaard N, Rasmussen S, et al. Structural insights into the regulatory mechanism of the response regulator RocR from Pseudomonas aeruginosa in cyclic di-GMP signalling. J Bacteriol. 2012;194:4837–46.

78. Christen M, Christen B, Folcher M, Schaar T, Jentsch T, et al. Identification of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GDP. J Biol Chem. 2002;277:30820–37.

79. Newell PD, Monds RD, Oyena K, Jiang W, Zhao M, Ling J, et al. A systematic analysis of the basis underlying functional diversification of cyclic di-nucleotide turnover proteins. J Mol Biol. 2010;402:524–38. https://doi.org/10.1016/j.jmb.2010.07.050.