Comparative genomics analysis of c-di-GMP metabolism and regulation in *Microcystis aeruginosa*

Meng Chen¹, Chun-Yang Xu¹, Xu Wang¹, Chong-Yang Ren¹, Jiao Ding¹ and Li Li¹,²*

**Abstract**

**Background:** Cyanobacteria are of special concern because they proliferate in eutrophic water bodies worldwide and affect water quality. As an ancient photosynthetic microorganism, cyanobacteria can survive in ecologically diverse habitats because of their capacity to rapidly respond to environmental changes through a web of complex signaling networks, including using second messengers to regulate physiology or metabolism. A ubiquitous second messenger, bis-(3′,5′)-cyclic-dimeric-guanosine monophosphate (c-di-GMP), has been found to regulate essential behaviors in a few cyanobacteria but not *Microcystis*, which are the most dominant species in cyanobacterial blooms. In this study, comparative genomics analysis was performed to explore the genomic basis of c-di-GMP signaling in *Microcystis aeruginosa*.

**Results:** Proteins involved in c-di-GMP metabolism and regulation, such as diguanylate cyclases, phosphodiesterases, and PilZ-containing proteins, were encoded in *M. aeruginosa* genomes. However, the number of identified protein domains involved in c-di-GMP signaling was not proportional to the size of *M. aeruginosa* genomes (4.97 Mb in average). Pan-genome analysis showed that genes involved in c-di-GMP metabolism and regulation are conservative in *M. aeruginosa* strains. Phylogenetic analysis showed good congruence between the two types of phylogenetic trees based on 31 highly conserved protein-coding genes and sensor domain-coding genes. Propensity for gene loss analysis revealed that most of genes involved in c-di-GMP signaling are stable in *M. aeruginosa* strains. Moreover, bioinformatics and structure analysis of c-di-GMP signal-related GGDEF and EAL domains revealed that they all possess essential conserved amino acid residues that bind the substrate. In addition, it was also found that all selected *M. aeruginosa* genomes encode PilZ domain containing proteins.

**Conclusions:** Comparative genomics analysis of c-di-GMP metabolism and regulation in *M. aeruginosa* strains helped elucidating the genetic basis of c-di-GMP signaling pathways in *M. aeruginosa*. Knowledge of c-di-GMP metabolism and relevant signal regulatory processes in cyanobacteria can enhance our understanding of their adaptability to various environments and bloom-forming mechanism.

**Keywords:** *Microcystis aeruginosa*, Comparative genomics, C-di-GMP, Phylogenetic analysis, GGDEF, EAL, HD-GYP, PilZ

* Correspondence: lili@sdu.edu.cn

¹Shandong Provincial Key Laboratory of Water Pollution Control and Resource Reuse, School of Environmental Science and Engineering, Shandong University, Qingdao, China

²Shandong Provincial Engineering Center on Environmental Science and Technology, Jinan, China

© The Author(s). 2020 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Background

Cyanobacteria, which are phototrophic bacteria that survive in ecologically diverse habitats, have received growing attention because they have been forming toxic blooms in eutrophic water bodies worldwide for decades [1, 2]. Dense blooms are considered seriously harmful to aquatic ecosystems because of their deleterious effects on water quality, such as increased turbidity, smothering submerged aquatic vegetation, and producing taste and odor compounds [3, 4]. Moreover, some cyanobacteria species can synthesize toxic secondary metabolites, such as hepatotoxin microcystins that can inhibit eukaryotic protein phosphatases; thus, they threaten the function of water bodies for drinking, bathing, and fishing, and they also ultimately pose potential risks to animal and human health [5–7]. Cyanobacteria are able to inhabit most of Earth’s environments because they evolved mechanisms to monitor and rapidly adapt to environmental changes through a web of complex signaling networks, such as using second messengers to regulate physiology or metabolism [8].

Cyanobacteria must cope with variations in the external environment. They rely on signaling molecules to translate these changes into intracellular responses and mediate adaptation to ambient conditions. Once bacterial cells sense an external stimulus, such as light and temperature, the intracellular level of a second messenger rapidly changes to amplify the biological input signal to a downstream output effector and initiate physiological changes, including sugar metabolism, motility, and biofilm production [8, 9]. A ubiquitous second messenger, bis-(3’5’)-cyclic-dimeric-guanosine monophosphate (c-di-GMP), which was first identified as an allosteric activator of cellulose synthase in Glacocococcus xylinus in 1987, plays an important role in regulating biofilm formation or dispersal in response to various environmental cues and cell–cell signals [10–14]. Studies have summarized that c-di-GMP regulates an astounding array of important processes in bacteria, including transcription, RNA turnover, protein synthesis, motility, virulence, and altering activities of proteins or protein complexes [15–17]. The intracellular level of c-di-GMP are modified by the rate of its synthesis and degradation in response to a variety of environmental stimuli, relying on the opposite enzymatic activity of diguanylate cyclases (DGCs) and c-di-GMP-specific phosphodiesterases (PDEs), respectively [12, 18]. DGC proteins contain a GGDEF domain that synthesizes one c-di-GMP molecule from two GTP molecules [19, 20]. PDE proteins contain an EAL or, less frequently, a HD-GYP domain, which breaks down c-di-GMP into the linear molecule 5’-phosphoguananylyl-(3’5’)-guanosine (pGpG) or into two GMP molecules [21, 22]. Moreover, GGDEF and EAL domains can both be present in the same protein, forming “hybrid” proteins, even though they have opposing activities [23, 24]. In that case, only one of the two domains is catalytically active, and the other performs a regulatory function, or a third regulatory domain is present that may disjoin the activity of the GGDEF and EAL domains [23, 25, 26]. Ute Römling et al. list a census of all GGDEF, EAL, and HD-GYP domains in bacterial genomes [12, 27].

Diverse sensor domains can modulate enzymatic activities in response to external stimuli, including N-terminal response regulator receiver (REC), Per/Arnt/Sim (PAS), histidine kinases/adenylate cyclases/methyl accepting proteins and phosphatases (HAMP), and cGMP phosphodiesterase/adenyl cyclase/FhlA (GAF) domains, which were related to c-di-GMP association network retrieved by STRING [25, 28–30]. C-di-GMP has been found to be recognized by downstream receptors that have been linked to specific physiological processes, ranging from polysaccharide biosynthesis to direct regulation of gene expression and to motility. Among the downstream effectors, the PilZ domain is ubiquitous in bacteria and can bind c-di-GMP to regulate biosynthesis of biofilms, such as cellulose and alginate [31–33]. The PilZ domain can be a stand-alone protein or fused with other functional proteins, such as cellulose synthases and alginate biosynthesis protein, or attached to certain signaling domains, such as the GGDEF, EAL, and HD-GYP domains [32, 34](Fig. 1). Molecular mechanisms of c-di-GMP signaling in a few cyanobacteria that are obligate photosynthetic microorganisms in the environment, such as Thermosynechococcus and Synechocystis, have been examined in-depth [35–38]. However, none of those studies have addressed Microcystis, one of the most ubiquitous freshwater cyanobacterial genera, which limits the comprehensive understanding of c-di-GMP signaling in cyanobacteria [36].
Results

General genome features of *M. aeruginosa* strains

Genomes of 24 *M. aeruginosa* strains were retrieved from the National Center for Biotechnology Information (NCBI) database for series analysis. As shown in Table 1, except for strains NIES 2481 [39] and NIES 2549 [40], no plasmid sequence was discovered in other strains. Among them, genome sequences of the strains CHAOHU 1326 and NaRes975 were recently released by our laboratory [41, 42], and their general information are shown in Table S1. The average size of the genomes was 4.97 ± 0.40 Mb, and the average G + C content was 42.66 ± 0.26%. Genomes ranged in size from 4.26 Mb (*M. aeruginosa* PCC9806) to 5.89 Mb (*M. aeruginosa* KW, Table 1). The selected genomes are of high completeness and low contamination as evaluated based on lineage-specific marker sets by checkM [43]. Multiple rRNA coding sequences were present in *M. aeruginosa* strains. Generally, each strain contains 1–2 sets of rRNA clusters as a rough estimation due to the incomplete sequences present in the genomes (Additional file 1, Table S2).

Modular signaling proteins involved in c-di-GMP metabolism and regulation in *M. aeruginosa*

A genome search for genes that encode enzymes involved in c-di-GMP metabolism was performed to identify the putative translated products that have DGC and PDE activities in the selected 24 *M. aeruginosa* genomes. The accession numbers of the predicted proteins are shown in Table 2. This survey led to identification of three enzymatic classes of predicted proteins DGCs, PDEs, and hybrid DGC–PDEs, which contain GGDEF domain and PDE proteins bearing the EAL or HD-GYP domains, respectively. Downstream effectors contain PilZ, degenerate GGDEF or EAL domain, riboswitch, and transcription factors.

---

**Fig. 1** C-di-GMP metabolism and regulation module. Once bacterial cells sense environmental cues, such as light and temperature, the intracellular level of c-di-GMP rapidly changes to amplify the biological input signal. C-di-GMP then is recognized by downstream effector that interact with a downstream target to affect specific physiological processes. C-di-GMP synthesis and degradation is achieved by DGC proteins containing the GGDEF domain and PDE proteins bearing the EAL or HD-GYP domains, respectively. Downstream effectors contain PilZ, degenerate GGDEF or EAL domain, riboswitch, and transcription factors.
DICT domain, a sensory domain in “diguanylate cyclases and two-component system”. Interestingly, compared with the HD-GYP domain-containing PDEs, which were identified in all selected *M. aeruginosa* genomes and seemed to be highly conserved, proteins with partial EAL domains were found less frequently (in only three
genomes). Except for the NIES44 genome, each of the other 23 genomes was found to have a GG [D/E]EF-EAL hybrid protein, consisting of GGDEF, EAL domain, and Forkhead associated (FHA) domain, a putative nuclear signaling domain.

Bacteria encode a variety of sensory and signal transduction proteins to sense and adapt to changes in the physicochemical makeup of their environment. Sensory and signal transduction proteins encoded in the selected 24 M. aeruginosa genomes were predicted, and 12 sensory domain-containing proteins were found. Most of these proteins are signal transduction histidine kinases. The accession numbers and domain architectures of the highly conserved GAF, PAS, and REC domain-containing proteins are listed in Additional file 1, Table S3. As an important sensor for photosensory behavior, the GAF domain was commonly associated with c-di-GMP domains in cyanobacteria [38]. As many as 11 of the 12 proteins had the GAF domain, and some even contained two. PAS-containing proteins are related to sensory input (GAF), transduction (HAMP), or output (histidine kinases). Half of the four predicted PAS-containing proteins contain a PAC motif, a conserved region of 40–45 amino acids located at the carboxy-terminal of the PAS domain, which contributes to PAS structure [28]. Interestingly, some sensory domain-containing proteins in different genomes were identical, and were therefore assigned the same accession number, such as NIES2549 and NIES2481, DIANCHI905 and PCC7806SL, and NaRes975 and PCC9808.

**Pan-genome of M. aeruginosa**

To assess the distribution of genes involved in c-di-GMP metabolism and regulation across the M. aeruginosa genome, a core–pan-genome analysis was performed using all 24 M. aeruginosa genome sequences as input in the Bacterial Pan Genome Analysis (BPGA) tool [44]. The pan-genome analysis revealed a core genome of 1918 genes with an accessory genome of 36,550 genes and 6489 unique genes (Fig. 2a). Accessory genes are those whose orthologs are present in two or more genomes, but not in all the genomes. M. aeruginosa possess a core genome shared by 24 strains, accounting for

---

**Table 2** Predicted modular signaling proteins involved in c-di-GMP metabolism in all 24 analyzed M. aeruginosa genomes

| Strains   | DGC (REC-GGDEF)* | PDE (DICT-HD-GYP) | PDE (EAL) | Hybrid protein (FHA-GGDEF-EAL) | DGC, PDE, Hybrid protein [a] |
|-----------|------------------|------------------|-----------|-------------------------------|-------------------------------|
| CHAOHU 1326 | WP_052276147.1 | WP_052275339.1 | – | WP_052277914.1 | 1, 1, 1 |
| DIANCHI905 | – | WP_002746813.1 | – | WP_002743531.1 | 0, 1, 1 |
| KW | WP_079210059.1 | WP_002796380.1 | – | WP_079210289.1 | 1, 1, 1 |
| NaRes975 | – | WP_002752229.1 | – | WP_044034220.1 | 0, 1, 1 |
| NIES44 | – | WP_045358386.1 | – | – | 0, 1, 0 |
| NIES87 | – | WP_104396273.1 | – | WP_104397223.1 | 0, 1, 1 |
| NIES88 | WP_061433230.1 | WP_061432432.1 | – | WP_061431785.1 | 1, 1, 1 |
| NIES98 | – | WP_002752229.1 | – | WP_002739484.1 | 0, 1, 1 |
| NIES843 | WP_012264732.1 | WP_002796380.1 | – | WP_012266621.1 | 1, 1, 1 |
| NIES1211 | WP_039900524.1 | WP_039900517.1 | WP_071989022.1 | WP_110544382.1 | 1, 1, 1 |
| NIES2481 | WP_046660716.1 | WP_06602931.1 | WP_080949698.1 | WP_066029445.1 | 1, 1, 1 |
| NIES2549 | WP_046660716.1 | WP_046662116.1 | WP_080949698.1 | WP_046660636.1 | 1, 1, 1 |
| PCC7806SL | – | WP_002746813.1 | – | WP_002743531.1 | 0, 1, 1 |
| PCC7941 | – | WP_002752229.1 | – | WP_043997363.1 | 0, 1, 1 |
| PCC9432 | – | WP_002752229.1 | – | WP_002750015.1 | 0, 1, 1 |
| PCC9443 | WP_043996837.1 | WP_002765696.1 | – | WP_002768060.1 | 1, 1, 1 |
| PCC9701 | WP_002801860.1 | WP_002803155.1 | – | WP_004163835.1 | 1, 1, 1 |
| PCC9717 | WP_043999403.1 | WP_002762031.1 | – | WP_002761714.1 | 1, 1, 1 |
| PCC9806 | WP_002783698.1 | WP_002780038.1 | – | WP_002783280.1 | 1, 1, 1 |
| PCC9807 | WP_002787322.1 | WP_002785975.1 | – | WP_004161732.1 | 1, 1, 1 |
| PCC9808 | – | WP_002752229.1 | – | WP_044034220.1 | 0, 1, 1 |
| PCC9809 | WP_043999403.1 | WP_002796380.1 | – | WP_002797049.1 | 1, 1, 1 |
| Sj | WP_110579156.1 | WP_110579081.1 | – | WP_110577728.1 | 1, 1, 1 |
| TAIHU98 | – | WP_002733640.1 | – | WP_002739484.1 | 0, 1, 1 |

*a Letters in parentheses are domains of the referred c-di-GMP metabolism enzymes

b Number of DGCs, PDEs and hybrids proteins
37.0 to 49.9% of the genome repertoire. The core–pan plot (Fig. 2b) showed that the pan-genome trend curve did not reach a plateau and seemed to extend with addition of more genomes to the analysis. Therefore, the pan-genome was considered an “open” pan-genome. In contrast, as shown in Fig. 2b, the core genome curve leveled off, considered as a conserved core genome.

The distribution of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Cluster of Orthologous Groups (COG) categories in *M. aeruginosa* has been mapped in Fig. 2c and Additional file 1, Figure S1. Genes involved in c-di-GMP metabolism and regulation were assigned to gene families belonging to environmental information and signal transduction category. Among these genes, the GG(D/E)EF domain-containing DGCs and GG(D/E)EF-EAL hybrid proteins coding genes were classed into accessory genes, while the HD-GYP domain containing PDEs and the sensory genes which possess
essential sensor and regulator domain, as the core genes, were relatively conservative for *M. aeruginosa* genes.

The BLAST Ring Image Generator (BRIG) alignment made it clear that most regions within the 24 *M. aeruginosa* genomes were conserved when compared to the reference strain NIES843 (Fig. 3). Several regions appeared to have low or even no similarity, possibly because of acquisition/deletion/rearrangement or horizontal gene transfer (HGT). The outer ring in Fig. 3 showed the distribution of genes encoding the domains involved in c-di-GMP signaling in NIES843 genome, and the specific locations were listed at Additional file 1, Table S4. The corresponding sequences of other *M. aeruginosa* strains seemed highly conserved, and some even shared an identity up to 100% with the reference genome.

**Phylogenetic analysis of *M. aeruginosa* strains**

Comparing with the phylogenetic tree based on the 16S rRNA gene (Additional file 1, Figure S2a), the phylogenetic tree based on the conserved marker genes, previously validated as phylogenetic markers for (cyano)bacteria [45], produced higher resolution (Fig. S2b). The 31 conserved marker genes tree revealed a topology with generally well-defined nodes, with bootstrap support values greater than 90% over 1000 replicates. Further, propensity for gene loss (PGL) analysis of the gene families revealed a group of strains have lost the REC-GGDEF domain coding gene, including strains NIES98 [46], TAIHU98, NaRes975, PCC9808, PCC9432, DIANCHI905, PCC7806SL, PCC7941, NIES87, and NIES44 (Fig. 4a). As to the node, consisting of strains NIES1211, PCC9701, NIES44, NIES2549 and NIES2481, the EAL domain coding gene seems to be acquired, but PCC9701 and NIES44 have lost this gene. Gene encoding GGDEF-EAL hybrid protein was lost in strain NIES44. There is no gain or loss of genes encoding HD-GYP and sensory domain-containing protein. To further analyze the evolution of genes encoding sensory domain-containing protein, phylogenetic tree was constructed using a multilocus sequence typing approach based on these concatenated conserved sensory domain-containing proteins sequences (Fig. 4b). It showed a similar topology.
with the conserved marker genes tree (Fig. S2b). Phylogenomic analyses based on binary gene presence/absence (1/0) pan-genome matrix generated by BPGA pipeline resulted in a tree (Fig. 4c) with a topology similar to the trees obtained using conserved marker genes and sensory genes (Figs. S2b, 4b). All phylogenetic trees provided more robust topologies than that based on 16S rRNA gene analysis alone.

A visual comparison of phylogenies based on 31 marker genes, sensory domains, and pan-genome presence/absence matrix were generated by tanglegram [47]. As shown in Fig. S3, only a few strains (2 of 24) occupied divergent positions on the phylogenetic trees based on 31 marker genes and sensory genes (Figs. S2b, 4b). All phylogenetic trees provided more robust topologies than that based on 16S rRNA gene analysis alone.

A visual comparison of phylogenies based on 31 marker genes, sensory domains, and pan-genome presence/absence matrix were generated by tanglegram [47]. As shown in Fig. S3, only a few strains (2 of 24) occupied divergent positions on the phylogenetic trees based on 31 marker genes and sensory genes, which indicated a congruence between the two trees (Fig. S3b). Interestingly, most of those strains in which REC-GGDEF domains containing protein DGC is not detected (marked in blue), including NIES98, TAIHU98, NaRes975, PCC9808, PCC9432, DIANCHI905, PCC7806SL, and PCC7941, appeared to be phylogenetically closely related, thus were always grouped in the same clade of the different phylogenetic trees based on 31 marker genes, sensory domains, and pan-genome presence/absence matrix. Among them, DIANCHI905 and PCC7806SL [48] are representatives of toxic (microcystin-producing) bloom-forming strains; in contrast, PCC 9432 and NIES98 are non-microcystin-producing strains [49]. Specifically, pairs of strains also appeared to be phylogenetically closely related, such as NIES2549 and NIES2481, DIANCHI905 and PCC7806SL, and NaRes975 and PCC9808, although they were isolated from diverse geographic origin (Table 1). The majority of the M. aeruginosa strains were isolated in different locations, but no correlation was found between their geographic distribution or bloom-forming ability and phylogenetic relationships, consistent with the previous report by Meyer et al [50].

To the majority sequence of genes encoding for c-di-GMP metabolism and regulation in M. aeruginosa, likelihood ratio test indicated that model M2 and M8 gave a significantly better fit than model M1 and M7, respectively, which allowed individual sites to evolve under positive selection (Additional file 1, Table S5). Lots of sites with $\omega > 1$ were identified in the sequence of genes responsible for c-di-GMP metabolism and regulation, revealing that they are likely to have been subjected to positive selection (Additional file 2, Table S6).

Structural features of GGDEF domain, EAL domain, and HD-GYP domain of M. aeruginosa strains

To elucidate the structural features, structure predictive modeling of GGDEF domain, EAL domain, and HD-GYP domain was performed on the corresponding M. aeruginosa strains. The NIES843 genome is a representative genome of M. aeruginosa because of its genome has been completely sequenced and is modeled in Fig. 5. Similarly, the corresponding structural models of strain CHAOHU 1326 were shown in Additional file 1, Figure S4. The Z-values of the models are shown in Table S7.
All the models of the GGDEF, EAL and HD-GYP domain containing proteins were qualified with a Z-score higher than −4.0.

According to SWISS-MODEL, the crystal structure of the conserved GGDEF domain of WspR (Protein Data Bank (PDB) id: 3BRE) was selected as the template to model the structure of the GGDEF domain of the DGC [51, 52]. Amino acids S173 to N329 from the GGDEF domain were used to perform structural alignments (Fig. 5a, left). The amino acid sequences of the GGDEF domains in DGC showed a similarity of 34.2–37.2% to that of 3BRE (Additional file 1, Table S8). C-di-GMP binds to the catalytic site and to a second site distal to the catalytic loop. DGC proteins possess a conserved allosteric inhibition site (I site),
composed of a RXXD motif (in which X represents any amino acid) five amino acids upstream of the GGDEF active site, that is important for controlling DGC activity. When levels of c-di-GMP are high, the second messenger can bind the RXXD motif, thereby repressing the DGC activity [53]. A systematic analysis and comparison of the 14 genomes that have corresponding GGDEF domains was performed to identify the amino acid motifs or signatures involved in catalysis and allosteric inhibition. As shown in Fig. 5a (left), the WebLogo alignment revealed that the RXXD and GGEDEF motifs of the GGEDEF domain were highly conserved in the same amino acid residues: Arg-Gln-Val-Asp (RQVD) and Gly-Gly-Glu-Glu-Phe (GGEDEF), respectively. The GG[DF/E]EF domain of the putative DGCs possessed the conserved amino acid residues essential for GTP binding, indicating that the DGCs may have catalytic activity [26].

Because only three genomes had partial EAL domains, the EAL domain in hybrid proteins from the *M. aeruginosa* NIES843 genome were chosen as paradigms to examine the crystal structure. Based on the crystal structure of the GGDEF-EAL domain of RmcA (PDB ID: 5M3C), which has a crystallographic resolution of 2.8 Å [54], the GGDEF and EAL domains in the hybrid protein of NIES843 were modeled. Compared with 5M3C, the GGDEF-EAL domains in the hybrid proteins showed sequence conservation of 35.9–37.8% (Additional file 1, Table S9). The low sequence conservation appeared to have no impact on model prediction by SWISS-MODEL. Compared with DGCs that contained only the GGDEF domain, amino acid residues of RXXD and GGDEF motifs in the GGDEF domain of the hybrid proteins were less conserved (Fig. 5a, right). The WebLogo alignment in Fig. 5b showed that amino acid residues of the EAL domain involved in the binding of c-di-GMP and catalytic activity were highly conserved in all sequences. The Glu in the EAL signature motif is an essential residue that is required to bind the c-di-GMP, whereas a change of Ala into Val (EVL) still sustains the enzymatic activity [55]. Arg in the second position downstream of the EAL signature motif was conserved in nearly all EAL domain sequences; thus, the EAL signature motif can be extended as EAXXR motif, which forms a stable platform to bind c-di-GMP [23].

Crystal structure of HD-GYP domain of *M. aeruginosa* NIES843 was modeled based on PA4781(PDB ID: 4R8Z) from *P. aeruginosa* [56] (Fig. 5c). Aligned with PA4781, the HD-GYP domain containing PDEs in *M. aeruginosa* showed sequence conservation of 33.5–34.3% (Additional file 1, Table S10). Generally, in HD-GYP domain, the HD residues clearly serve as metal ligands, the signature of HD can be extended as a larger motifs HDxGK; while the GYP motif may be more usefully considered as part of the HHExxDxGxGYP, and the role of GYP motif may be substrate specificity determining but is not certainly clear [56, 57]. WebLogo alignment revealed that the GYP motifs of GYP loop in *M. aeruginosa* were highly conserved in the same amino acid residues EFExxDxGxGVP, whereas Val replaced Tyr compared with the GYP motif template. Moreover, the HD motif possess YR residues in *M. aeruginosa* strains. That is, amino acid residues were replaced by YR-GVP in HD-GYP motif in all selected *M. aeruginosa* strains.

**Structural features of the PilZ domain of *M. aeruginosa* strains**

All selected *M. aeruginosa* genomes encoded proteins that possess a PilZ domain. Twenty genomes encoded cellulose synthase (CelA), which contained a C-terminal PilZ domain, and the other four genomes encoded a protein that contained only a PilZ domain. The accession numbers of the corresponding proteins are shown in Additional file 1, Table S11.

To identify the structural features, structure predictive modeling of proteins with a single PilZ domain and PilZ domain-containing CelA was performed for *M. aeruginosa* strains. Predictive modeling was based on the crystal structure of the BcsA (PDB id: 4P02) from *Rhodobacter sphaeroides*, which has a crystallographic resolution of 2.65 Å, according to SWISS-MODEL results [31]. Model of the PilZ domain-containing protein CelA of strain CHAOHU 1326 is shown in Fig. 6a. The c-di-GMP-binding PilZ domain was located in the C-terminal region of CelA and had similar structure with protein containing a single PilZ domain in Fig. 6b, which was derived from the representative *M. aeruginosa* strain NIES843. Figure 6c shows that the PilZ domain consists of a six-stranded β-barrel and a short α-helix that follows the last strand of the β-barrel.

**Discussion**

The occurrence of cyanobacterial blooms appears to be increasing because of environmental factors, including continued eutrophication, rising atmospheric CO₂ concentrations, and global warming [58–60]. Cyanobacteria can survive in ecologically diverse habitats, to a great extent, because intracellular second messengers function in pathways that mediate cellular responses to oxidative stress, nutrient imbalances, and temperature variations in the environment [8]. c-di-GMP, as a universal bacterial second messenger, has been shown to regulate biofilm formation and aggregation, which are beneficial for cyanobacteria colony formation and thus promotes bloom formation [35, 61]. With recent advances in genome sequencing and bioinformatics, it is possible to identify sequence groups with high genotypic similarity based on variation in protein-coding genes distributed across the genomes and predictions drawn from bioinformatics, and thereby provide genetic insight into c-di-GMP signaling regulation in *M. aeruginosa*. Because only one or two *M.
aeruginosa genomes do not adequately represent this species, 24 M. aeruginosa genomes available in NCBI’s GenBank were selected to comprehensively clarify the genetic similarities and differences of M. aeruginosa strains in the present study.

The selected M. aeruginosa strains in this study diverged to some extent at the genomic level and were isolated from aquatic ecosystems around the world. An in-depth comparative genomics analysis that included genome feature analysis, core-pan-genome analysis, and phylogenetic analysis were used to distinguish differences and similarities among the 24 selected M. aeruginosa genomes.

M. aeruginosa genome sizes result from a mix of gains and losses during natural selection as they were subjected to changing environments and competitive forces during the evolution of the species. As a freshwater species, M. aeruginosa have medium genomes compared with other cyanobacteria, especially compared with marine species that mostly occur in low nutrient and stable open ocean waters, such as Synechococcus and Prochlorococcus, the genome sizes of which are almost half those of M. aeruginosa [62]. Some reports indicated that genome size is positively correlated with the number of duplicated genes, which can originate from either within the genome itself or can be introduced by HGT [63]. Gene duplication and high genetic redundancy in the M. aeruginosa genomes are considered an evolutionary strategy that might confer this cyanobacterial species an extensive adaptive capacity that allows them to inhabit a wide range of habitats worldwide, and facilitates the ability to proliferate and dominate the phytoplankton communities in eutrophic freshwater ecosystems [64]. The core-pan-genome analysis indicated that these strains maintained a conserved core genome and an expansive pan-genome that allow them to acquire new genes. M. aeruginosa possess a relatively small core genome, which might result from high genetic diversity and variable gene content [49].

In this study, bioinformatics tools furthered our understanding of c-di-GMP signaling in M. aeruginosa by recognizing and studying domain architectures and tridimensional structures of the predicted proteins with DCGs, PDEs, and DGC–PDEs encoded in the genomes. These coding genes are widespread in other cyanobacterial species, such as Synechocystis sp. PCC6803 and Thermosynechococcus elongatus BP-1, which reportedly encode a considerable number of proteins predicted to be involved in c-di-GMP metabolism [36, 38]. In general, the number of domains involved in c-di-GMP signaling in cyanobacteria may be mainly determined by genome size [65]. However, there are at most three c-di-GMP signal-related domains identified in M. aeruginosa genomes, even if the mean genome size of this species is nearly two-fold that of the Synechocystis sp. PCC6803 and Thermosynechococcus elongatus BP-1. An alternate explanation is that, in cyanobacteria, the number of c-di-GMP signal-related domains are not simply correlated with genome size but may also be affected by bacterial adaptation. Among the species present in the Cyanobase database, the species found to lack c-di-GMP
signaling systems were *Prochlorococcus* and some *Synechococcus* strains. It was reported that *Synechococcus* strains that contain c-di-GMP-modulating domains inhabit both marine and freshwater habitats and are found in rich-nutrient (eutrophic) waters, whereas *Synechococcus* strains lacking c-di-GMP-regulatory domains inhabit low-nutrient (oligotrophic) marine habitats [36]. Species adapted to stable habitats may have lost genes that encode c-di-GMP-modulating proteins. Primitive *M. aeruginosa* that inhabit low-nutrient lakes may have a small number of c-di-GMP domains even though they have relatively large genomes [66].

The domain architectures of the deduced amino acid sequences from the *M. aeruginosa* genomes also revealed diverse sensor domains, such as REC, PAS/PAC, GAF, and HAMP, which are involved in activity regulation by driving the protein dimerization process and play important roles in c-di-GMP-controlled rapid response to changing environmental conditions. It seemed that these sensor domains have been subjected to positive selection during evolution. Some sensory domain-containing proteins from different genomes have identical amino acid sequences, such as that of NIES2549 and NIES2481, DIANCHI905 and PCC7806SL, and NaRes975 and PCC9808. It should be noted that each pair of strains have close genetic relationships as determined by phylogenetic analysis.

In this study, four types of phylogenetic trees were established based on 16S rRNA gene, 31 protein-coding phylogenetic marker genes, sensory protein sequences, and binary gene presence/absence (1/0) pan matrix. The congruence of the two phylogenetic trees based on 31 protein-coding phylogenetic marker genes and sensory protein sequences facilitated a comprehensive understanding of the phylogenetic relationships and the evolution of the sensory domain coding genes among *M. aeruginosa* strains. It should be noticed that most of strains in which RECGGDEF domains-containing protein DGC was not detected, appeared to be phylogenetically closely related. PGL analysis revealed most of genes involved in c-di-GMP signaling are stable in *M. aeruginosa* strains. In some *M. aeruginosa* strains, the EAL domain coding gene seems to be acquired from other origin by lateral gene transfer. In complex signal transduction process, the range of cellular functions might be regulated by different regulatory systems. The number of genes involved in c-di-GMP signaling might more likely rely on the number (kinds) of signals that they respond to and the intracellular level of the c-di-GMP that they need to regulate a series of physiological process, instead of genome size [67]. In some *M. aeruginosa* strains, the missing or lack of coding genes for DGC or PDE revealed that c-di-GMP-mediated regulation might not be the sole alternative regulatory pathway in this ancient photosynthetic microorganism. They might use other signal molecules such as NO, to regulate diverse biochemical and physiological processes [8].

It seems that the relatedness of the closely related strains studied did not perfectly reflect their similar physiological characteristics (e.g., cyanobacterial toxin-producing ability) or geographical origins. Phylogenetic analysis could also not reveal the *M. aeruginosa* strains with bloom-forming characteristics. Previous studies demonstrated that *Microcystis* “species” distinctions are problematic and doubtful [68, 69]. *Microcystis* taxonomic studies using 16S rRNA analysis revealed that phylogenetic trees using sequences with significantly high sequence similarities did not clearly delineate *Microcystis* species [70, 71]. In the present study, 31 protein-coding phylogenetic marker genes was used instead of discrimination based only on the traditional 16S rRNA gene, which could not sufficiently discriminate between strains [44, 45]. Similarly, the phylogenomic tree based on whole genome information was more reliable compared with the phylogenetic tree only based on the 16S rRNA gene. More tests are needed to further determine whether the alternative approaches could refine *Microcystis* species classification.

GG[D/E]EF and EAL domain-containing proteins analyzed in this study included all essential conserved amino acid residues that bind the corresponding substrate to have enzymatic activity. In *M. aeruginosa*, HD-GYP domains possess the variant key residues YR-GVP, and further study needs to be done to verify whether this domain still has the ability to bind corresponding substrates or is degenerate. Structural analysis provides important information for predicting the function of these proteins that contain GGDEF, EAL, hybrid domains, and HD-GYP domain, and creates a paradigm for future studies that analyze the evolution of enzymes involved in c-di-GMP metabolism. It was also found that all selected *M. aeruginosa* genomes encode PilZ domain, regardless of if it is in CelA, by which c-di-GMP could stimulate the biosynthesis of extracellular polysaccharides that are important for biofilm formation.

**Conclusion**

In summary, comparative genomic analysis of 24 publicly available *M. aeruginosa* genomes focusing on c-di-GMP metabolism and regulation revealed the following main results:

1. Proteins involved in c-di-GMP metabolism and regulation, such as diguanylate cyclases, phosphodiesterases, and PilZ-containing proteins, were encoded in *M. aeruginosa* genomes. However, the number of identified c-di-GMP signaling related domains was not proportional to the size of *M. aeruginosa* genomes (4.97 Mb in average). Pan-genome analysis showed that genes
involved in c-di-GMP metabolism and regulation are relatively conservative in *M. aeruginosa* strains. (2) Phylogenetic and phylogenic analysis revealed that the relatedness of the closely related *M. aeruginosa* strains did not reflect the geographical origins, even though they were isolated from diverse freshwater ecological environments. PGL analysis revealed that most of c-di-GMP signaling related genes are stable in *M. aeruginosa* strains. (3) In silico analysis of signaling related DGCs, PDEs, and hybrid proteins revealed that GGDEF and EAL domains contained the conserved amino acid residues essential for the substrates binding, indicating a possible catalytic activity. In addition, it was also found that all selected *M. aeruginosa* genomes encode PilZ domain, regardless of if it is in CelA.

This study is the first to analyze c-di-GMP signal-related proteins in *M. aeruginosa*, and our findings provide a pre-requisite genetic basis for further experimental characterization and evaluation of biological function. Some important aspects are still unclear that could help enhance our understanding of *M. aeruginosa* blooms in aquatic environments, such as the involvement of the specific domain-containing proteins of c-di-GMP signaling networks in *M. aeruginosa* physiological regulation and an ecologically relevant explanation of how *M. aeruginosa* adapts to its specific ecological niche.

**Methods**

*M. aeruginosa* genomes features

All of the *M. aeruginosa* genome sequences available in June 2018 in the NCBI database, annotated with the Prokaryotic Genome Annotation Pipeline [72], were used to conduct various analyses. Draft genomes that consisted of more than 1000 contigs were omitted to obtain consistent genome quality. CheckM v.1.0.8 was used to estimate the completeness of the selected *M. aeruginosa* genomes [43]. The genome features of these selected strains were listed in Table 1. The sequencing and sequence assembly of *M. aeruginosa* strain NaRes975 and CHAOHU 1326 genomes were performed as previously described [41].

Identification of genes involved in c-di-GMP metabolism and regulation

Genes that encode the GG[DE]EF, EAL, and GG[DE]EF-EAL domains, the related sensor GAF, PAS, and HAMP domains, and the c-di-GMP binding domain PilZ from the selected 24 *M. aeruginosa* genome sequences were identified by performing BLAST (Identity ≥30% for amino acid and 80% for nucleotide, E-value ≤1E−5). Conserved Domain Database (CDD) [73], Microbial Signal Transduction Database (MiST, version 3.0) [74], and Simple Modular Architecture Research Tool (SMART) [75] were used to characterize the domain.

**Comparative genome analyses**

Core–pan-genome analysis was performed using the BPGA tool [44]. Orthologous clusters were assigned by grouping all protein sequences encoded by the 24 genomes using the default clustering tool USEARCH based on 90% sequence identity cut-off. Core–pan-genome plots were calculated over 500 iterations. Comparative functional analysis was performed based on COG of proteins and KEGG pathways by focusing on distributions of representative protein sequences of core, accessory, and unique clusters of the *M. aeruginosa* strains. Gene families were classified accordingly. BRIG(version 0.95) [76] was used to create a circular genome comparison to highlight the location of genes encoding c-di-GMP-associated signaling domains between the 24 genomes compared with the reference sequence.

**Phylogeny and evolution analysis**

To elucidate the phylogenetic relationships between the *M. aeruginosa* strains, 16S rRNA gene sequences of cyanobacterial strains for which whole genome sequence data were available on NCBI were downloaded and analyzed to construct a phylogenetic tree. Sequences were aligned in MUSCLE version 3.8 with default settings [77]; then, the phylogenetic and molecular evolutionary analyses were conducted using MEGA version X [78]. The phylogenetic tree was inferred using the neighbor-joining method with 1000 bootstrap replications. The evolutionary distances were computed using the maximum composite likelihood method and the units were number of base substitutions per site. The analysis involved 25 nucleotide sequences, including 24 sequences of *M. aeruginosa* strains and the *Synechocystis* sp. PCC6803 sequence as the outgroup. All positions that contained gaps and missing data were eliminated. There were a total of 1313 nucleotides in the final dataset.

To assess relationships between the *M. aeruginosa* strains, the phylogenetic tree was constructed based on amino acid sequences of 31 highly conserved proteins that were encoded by the genes distributed in genomes as a single copy along 24 *M. aeruginosa* genomes, according Wu and Eisen (2008) [45, 79]. These protein sequences were mined by the AutoMated Phylogenomic inference Application–AMPHORA2 tool [80, 81], using default settings for the bacteria option and a cut-off E-value of 1 E-10. Individual alignments were performed for each of the 31 gene sets in MUSCLE version 3.8 with default settings [77], trimmed with respect to the reading frame, and subsequently concatenated with the FaBox Fasta Alignment Joiner [82]. Only genomes with all selected sets of conserved genes were used in the phylogenetic analysis. A Maximum Likelihood (ML) tree was constructed with MEGA X using the Jones–Taylor–
from *R. sphaeroides* resolution structures that were then used to create the for models whose scores strongly matched high-resolution structures that were ranked based on the expected quality of the result-structure domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aeruginosa* domain containing protein PA4781 from *P. aerugi-
Orthologous Groups; DGCs: Diguanylate cyclases; FHA: Forkhead associated; GAF: cGMP phosphodiesterase/adenylyl cyclase/FH/Na; HAMP: Histidine kinases/adenylate cyclases/methyl accepting proteins and phosphatases; HGT: Horizontal gene transfer; KEGG: Kyoto Encyclopedia of Genes and Genomes; MiST: Microbial Signal Transduction Database; ML: Maximum Likelihood; NCBI: National Center for Biotechnology Information; ncRNA: Noncoding RNA; PAS: Per/Arnt/Sim; PDB: Protein Data Bank; PDEs: Phosphodiesterases; PGL: Propensity for gene loss; pGpGC: 5′-phosphoguanosyl(3′-5′)-guanosine; REC: Response regulator receiver; SMART: Simple Modular Architecture Research Tool.

Acknowledgements
We thank Mallory Eckstut, PhD, from Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac), for editing the English text of a draft of this manuscript.

Authors’ contributions
LL conceived the study and supervised the research. MC received the sequences from the NCBI GenBank database under the accession numbers as provided in Table 1. Specifically, the whole genome sequence of M. aeruginosa CHAOHJU 1326 and NALR075 have been deposited in the GenBank database under accession number MOLZ0000000 and MOLN0000000, respectively. All data generated during this study are included within the paper and/or additional files.

Availability of data and materials
All genome sequences analyzed in the current study are available on the NCBI GenBank database under the accession numbers as provided in Table 1. Specifically, the whole genome sequence of M. aeruginosa CHAOHJU 1326 and NALR075 have been deposited in the GenBank database under accession number MOLZ0000000 and MOLN0000000, respectively. All data generated during this study are included within the paper and/or additional files.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 18 September 2019 Accepted: 19 February 2020
Published online: 03 March 2020

References
1. Harke MJ, Steffen MM, Gobler CJ, Otten TG, Wilhelm SW, Wood SA, et al. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. Harmful Algae. 2016;54:4–20.
2. García-Pichel F, Belnap J, Neuer S, Schanz F. Estimates of global cyanobacterial biomass and its distribution. Arch Hydrobiol Suppl Algal Stud. 2003;109:213–27. https://doi.org/10.1127/1864-1318/2003/0109-0213.
3. Parel HW, Huisman J. Blooms like it hot. Science. 2008;320(5872):57–60. https://doi.org/10.1126/science.1151631.
4. 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):1–52. https://doi.org/10.1128/MMBR.00043-12.
5. Sauer K. In: Walker JM, editor. c-di-GMP Signaling; 2017.
6. Boyd CD, O’Toole GA. Second messenger regulation of biofilm formation: breakthroughs in understanding c-di-GMP effector systems. Annu Rev Cell Dev Biol. 2012;28(28):439–62.
7. 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):1–52. https://doi.org/10.1128/MMBR.00043-12.
8. Liang ZX. The expanding roles of c-di-GMP in the biosynthesis of exopolysaccharides and secondary metabolites. Nat Prod Rep. 2015;32(5):663–83. https://doi.org/10.1039/c4np00086b.
9. Ryjenkov DA, Tarutina M, Moskvin OV, Galperin MY, Gomelsky M. Cyclic diguanylate cyclase is a ubiquitous signaling molecule in bacteria insights into biochemistry of the GGDEF protein domain. J Bacteriol. 2005;187(5):1792–8. https://doi.org/10.1128/8.1792.1798.2005.
10. Chan C, Paul R, Samoray D, Amiot NC, Giese B, Jenal U, et al. Structural basis of activity and allosteric control of diguanylate cyclase. Proc Natl Acad Sci U S A. 2004;101(19):7084–9. https://doi.org/10.1073/pnas.0406141101.
11. Whiteley CG, Lee DJ. Bacterial diguanylate cyclases: structure, function and mechanism in exopolysaccharide biofilm development. Biotechnol Adv. 2015;33(1):124–41. https://doi.org/10.1016/j.biotechadv.2014.11.010.
12. Sultan SZ, Pirzer E, Boquill T, Hobbs G, Miller MR, Motaleb MA. Analysis of the HD-GYP domain cyclic dimeric GMP phosphodiesterase reveals a role in motility and the enzootic life cycle of Borelia burgdorferi. Infect Immun. 2011;79(8):2737–83. https://doi.org/10.1128/iai.05153-11.
13. 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(35):30829–37. https://doi.org/10.1074/jbc.M504292200.
14. Chow SH, Galperin MY. Diversity of cyclic di-GMP-binding proteins and mechanisms. J Bacteriol. 2016;198(1):32–46. https://doi.org/10.1128/jb.00333-15.
15. Navarro M, De N, Bae N, Wang Q, Sondermann H. Structural analysis of the GGDEF-EAL domain-containing c-di-GMP receptor FimK. Structure. 2009; 1708:1104–16. https://doi.org/10.1016/j.str.2009.06.010.
16. Galperin MY. Diversity of structure and function of response regulator output domains. Curr Opin Microbiol. 2010;13(2):1250–9. https://doi.org/10.1016/j.mib.2010.01.005.
17. Mata AR, Pacheco OM, Cruz Pérez JF, Sáenz NM, Baca BE. In silico comparative analysis of GGDEF and EAL domain signaling proteins from the Azospirillum genomes. BMC Microbiol. 2018;18(1):20.
18. Römling U, Galperin MY, Gomelsky M. Distribution of GGDEF, EAL, HD-GYP and PEZ domains in bacterial genomes 2013. Available from: https://www.ncbi.nlm.nih.gov/Complete_Genomes/c-di-GMP.html, Updated 2016 Aug 31.
19. Henry JT, Crosson S. Ligand-binding PAS domains in a genomic, cellular, and structural context. Annu Rev Microbiol. 2011;65(1):361–86. https://doi.org/10.1146/annurev-micro-101809-151631.
20. Schimer T. C-di-GMP synthesis: structural aspects of evolution, catalysis and regulation. J Mol Biol. 2016;429(19):3683–701. https://doi.org/10.1016/j.jmb.2016.07.023.
21. Szkarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, 22. Agostoni M, Montgomery BL. Survival strategies in the aquatic and terrestrial world: the impact of second messengers on cyanobacterial processes. Life Basel. 2014;4(4, Sp. Iss. 5):745–69. https://doi.org/10.3390/ life4040745.
23. Townesley L, Yildiz FH. Temperature affects c-di-GMP signalling and biofilm formation in Vibrio cholerae. Environ Microbiol. 2015;17(11):4290–305. https://doi.org/10.1111/1462-2920.12799.
24. Valenti M, Filloux A. Biofilms and cyclic di-GMP (c-di-GMP) signalling: lessons from Pseudomonas aeruginosa and other bacteria. J Biol Chem. 2016;291(24):12547–55. https://doi.org/10.1074/jbc.R115.711507.
25. An SW, Wu JE, Zhang LH. Modulation of Pseudomonas aeruginosa biofilm dispersal by a cyclic-di-GMP phosphodiesterase with a putative hypoxia-sensing domain. Appl Environ Microbiol. 2010;76(20):6160–70. https://doi.org/10.1128/aem.00233-10.
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(1):1–52. https://doi.org/10.1128/MMBR.00043-12.
27. Henrissat B, Chandonia JM. SMART: Simple Modular Architecture Research Tool.
