PLoS GENETICS

c-di-GMP Turn-Over in *Clostridium difficile* Is Controlled by a Plethora of Diguanylate Cyclases and Phosphodiesterases

Eric Bordeleau[1,2], Louis-Charles Fortier[2,3], François Malouin[1,2], Vincent Burrus[1,2]*

1 Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, Canada, 2 Centre d'Étude et de Valorisation de la Diversité Microbienne (CEVDM), Université de Sherbrooke, Sherbrooke, Canada, 3 Département de Microbiologie et d'Infectiologie, Faculté de Médecine et Sciences de la Santé, Université de Sherbrooke, Sherbrooke, Canada

Abstract

Clostridium difficile infections have become a major healthcare concern in the last decade during which the emergence of new strains has underscored this bacterium's capacity to cause persistent epidemics. c-di-GMP is a bacterial second messenger regulating diverse bacterial phenotypes, notably motility and biofilm formation, in proteobacteria such as *Vibrio cholerae*, *Pseudomonas aeruginosa*, and *Salmonella*. c-di-GMP is synthesized by diguanylate cyclases (DGCs) that contain a conserved GGDEF domain. It is degraded by phosphodiesterases (PDEs) that contain either an EAL or an HD-GYP conserved domain. Very little is known about the role of c-di-GMP in the regulation of phenotypes of Gram-positive or fastidious bacteria. Herein, we exposed the main components of c-di-GMP signalling in 20 genomes of *C. difficile*, revealed their prevalence, and predicted their enzymatic activity. Ectopic expression of 31 of these conserved genes was carried out in *V. cholerae* to evaluate their effect on motility and biofilm formation, two well-characterized phenotype alterations associated with intracellular c-di-GMP variation in this bacterium. Most of the predicted DGCs and PDEs were found to be active in the *V. cholerae* model. Expression of truncated versions of CD0522, a protein with two GGDEF domains and one EAL domain, suggests that it can act alternatively as a DGC or a PDE. The activity of one purified DGC (CD1420) and one purified PDE (CD0757) was confirmed by *in vitro* enzymatic assays. GTP was shown to be important for the PDE activity of CD0757. Our results indicate that, in contrast to most Gram-positive bacteria including its closest relatives, *C. difficile* encodes a large assortment of functional DGCs and PDEs, revealing that c-di-GMP signalling is an important and well-conserved signal transduction system in this human pathogen.

Introduction

Clostridium difficile is a Gram-positive, anaerobic, spore-forming bacterium causing mild diarrhea to fulminant colitis in humans. Due to spreading of hypervirulent and high toxin-producing strains, *C. difficile* has caused in the last decade several epidemics in Europe and North America where it is now the leading cause of nosocomial diarrhea [1–3]. Its ability to sporulate allows this bacterium to remain dormant for years and survive to harsh conditions such as gastric acid after ingestion or the presence of oxygen in the environment. *C. difficile*-associated diseases are commonly associated with antibiotic usage, which creates a favorable niche for *C. difficile* to grow and cause infection in part by disrupting the gut microflora.

Bis-(3′-5′)-cyclic dimeric guanosine monophosphate (c-di-GMP) is a bacterial second messenger controlling diverse bacterial phenotypes mostly known to be involved in the transition from free-living, motile to biofilm lifestyle in Gram-negative bacteria [4,5]. c-di-GMP has also been shown to be involved in the development and cell cycle control in *Caulobacter crescentus* [6,7], and the modulation of virulence in several pathogens such as *Vibrio cholerae*, *Vibrio vulnificus*, *Bordetella pertussis* or *Pseudomonas aeruginosa* [8–11]. c-di-GMP is synthesized from 2 GTP molecules by enzymes named diguanylate cyclases (DGCs) that contain a GGDEF domain [12]. It is degraded respectively into pGpG or 2 GMP by phosphodiesterases (PDEs) that contain an EAL (PDEA) or a HD-GYP domain [13,14]. GGDEF domains were named by Hecht and Newton according to their conserved amino acid motif GG[D/E]EF [12]. EAL and HD-GYP domains are also named based on conserved amino motifs within these domains [13,14]. Whole genome analysis of a large number of bacterial species has revealed that the number of genes coding for enzymes involved in c-di-GMP turn-over varies widely between different species [15]. Genomes of proteobacteria generally encode a much wider array of such enzymes compared to those of Gram-positive bacteria. For instance, 66 genes coding for predicted enzymes involved in c-di-GMP turn-over are found in *Shewanella oneidensis*, 62 in *V. cholerae*, 41 in *Pseudomonas aeruginosa*, 29 in *Escherichia coli*, 6 in *Bacillus subtilis*

Author Summary

c-di-GMP is a bacterial intracellular signalling molecule regulating motility, biofilm formation, cell cycle control, or virulence in Gram-negative bacteria. The function and importance of this molecule still remain unknown in Gram-positive bacteria, even in important emerging pathogens such as *Clostridium difficile*, which causes from mild to deadly intestinal infections and has lately been on the rise in the healthcare setting and in the community. Here, we expose in the genomes of *C. difficile* strains a large number of conserved genes encoding proteins involved in the synthesis, degradation, or sensing of c-di-GMP, in contrast with most other Gram-positive bacteria including *C. difficile*'s closest relatives. We confirmed the activity of most of these well-conserved proteins in a microorganism for which typical behavior alterations associated with variation of the intracellular c-di-GMP pools are known. We further confirmed the c-di-GMP synthesis and degradation activities of two purified *C. difficile* proteins. Our results indicate that c-di-GMP signalling is important in the lifecycle of this pathogen. This finding is particularly exciting because the c-di-GMP signalling network could serve as a target for the development of new drugs against *C. difficile*-associated diseases that are commonly associated with antibiotic usage.

and 1 in *Staphylococcus aureus*. In fact, very little is known about c-di-GMP's input in the regulation of phenotypes within Gram-positive bacteria. GdpS, the sole predicted staphylococcal GGDEF domain-containing protein, positively regulates biofilm formation in both *S. aureus* and *S. epidermidis*, and expression of protein A, a major virulence factor in *S. aureus*. However, GdpS does not appear to be an active DGC *in vitro* and its C-terminal GGDEF domain is not involved in these two phenotypes [16,17].

Although the enzymes producing and degrading c-di-GMP share customary domains in all bacteria, when known, the downstream effectors and pathways regulating the different phenotypical responses are usually different. c-di-GMP-sensing proteins containing the characterized c-di-GMP-binding PilZ domain [18] and some non-PilZ proteins are known c-di-GMP binding receptors. In *E. coli* and related bacteria, the PilZ domain-containing protein YcgR was found to decrease motility by interacting with the flagellar motor to control its direction and rotation speed upon binding of c-di-GMP [19,20]. In *V. cholerae*, VpsT, a key transcription regulator that inversely regulates the expression of genes associated with motility and biofilm formation, was only recently found to be active following binding of c-di-GMP to its atypical receiver domain [21]. Recently, the first c-di-GMP-binding riboswitches (c-di-GMP-I) have been discovered in the genomes of *V. cholerae*, *C. difficile* and other bacteria [22]. Riboswitch Cd1 of *C. difficile* is located upstream of the large operon coding for the synthesis of the flagellum and was found to have an ''off'' switch action on transcription in an *in vitro* transcription assay and in β-galactosidase assays in *B. subtilis* [22]. Additionally, self-splicing of an unusual group I intron in *C. difficile* genome was found to be allosterically controlled by a c-di-GMP-II riboswitch aptamer, likely enabling the translation of a putative surface protein upon c-di-GMP binding [23]. Furthermore, 37 genes of *C. difficile* 630 that encode putative proteins containing GGDEF and/or EAL domains are available in the SignalCensus database [24]. Together, these observations suggest that c-di-GMP is a key signalling component in this emerging pathogen; yet studies on proteins regulating the intracellular c-di-GMP level, i.e. DGCs and PDEs, are still lacking.

The identification and characterization of enzymes producing or degrading c-di-GMP is a critical step to determine and understand the relevance of this second messenger in *C. difficile*'s lifecycle and virulence. In this study, we analyzed the prevalence and conservation of genes coding for putative DGCs and PDEAs in the genomes of 20 *C. difficile* isolates. Thirty-one conserved genes were assayed for their ability to encode functional DGC or PDEAs by evaluating the effect of their expression in *V. cholerae*. Most of these proteins conferred phenotypes that were consistent with their predicted function in our heterologous expression model. Our results indicate that, unlike the vast majority of Gram-positive bacteria including the Clostridiaceae, *C. difficile* regulates its intracellular c-di-GMP pools via a plethora of functional DGCs and PDEAs.
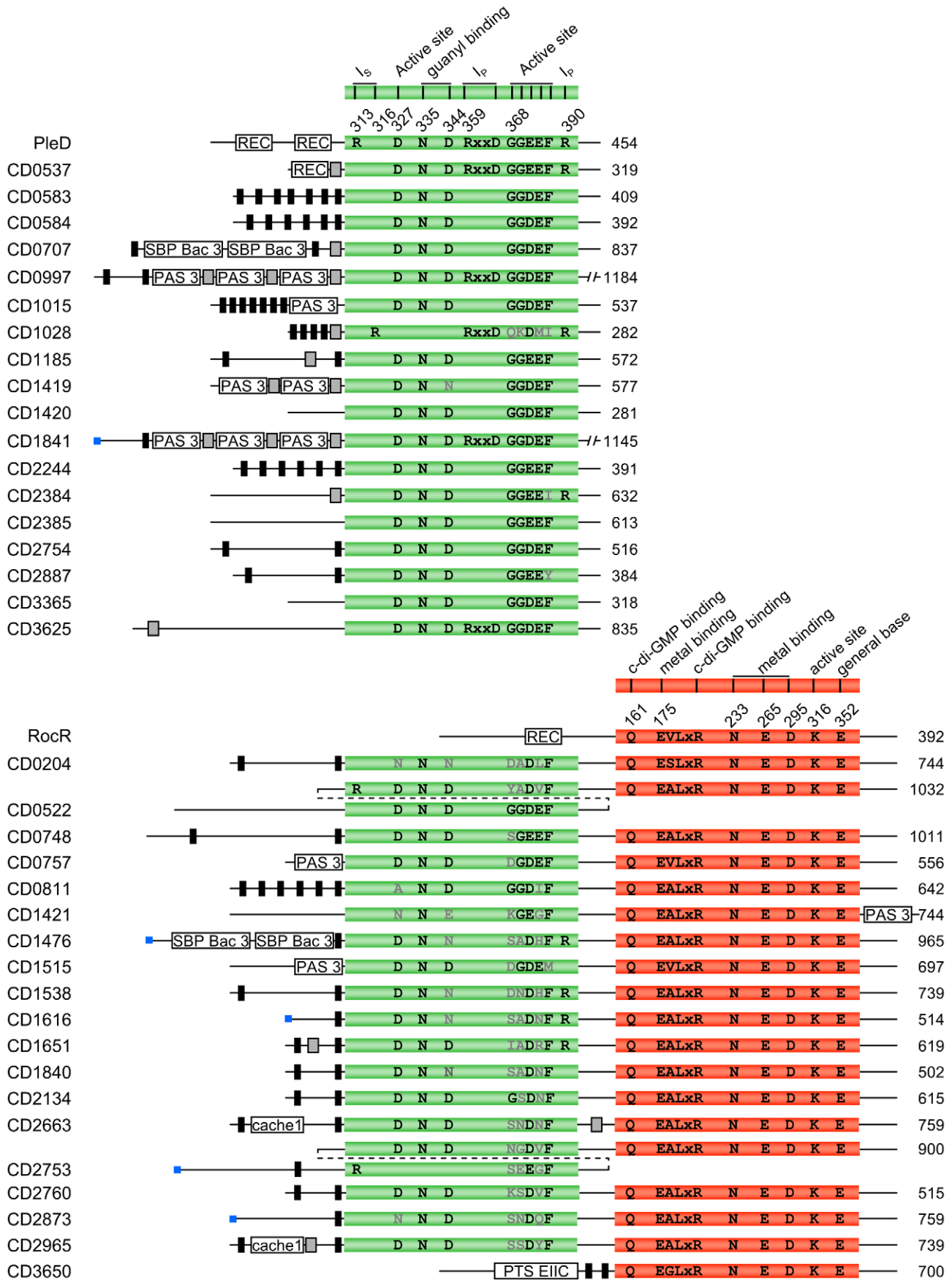
Results

Domain composition and activity prediction of putative DGCs and PDEAs encoded by *C. difficile* 630

Initial examination of the Pfam database 24.0 [25] for proteins involved in c-di-GMP turn-over in *C. difficile* 630, correlated to data provided by the NCBI's SignalCensus database [24], reveals a total of 37 proteins containing a GGDEF (Pfam PF00990) and/or an EAL (Pfam PF00563) domain: 18 proteins have a GGDEF domain, one protein contains an EAL domain and 18 proteins have both GGDEF and EAL domains. Proteins containing both GGDEF and EAL domains have been shown to act either as DGCs or as PDEAs, or to exhibit both activities [14,26,27]. Furthermore, not all proteins containing a GGDEF or an EAL domain have been shown to exhibit DGC or PDEA activity. Several conserved amino acids in the GGDEF and EAL domains are predicted to be important to confer enzymatic activity [28,29], among which the highly conserved motifs GG[D/E]EF and EXLR, respectively. These motifs and other conserved amino acid residues were sought for in the primary sequences of the 37 putative c-di-GMP-signalling proteins encoded by the genome of *C. difficile* 630 by multiple sequence alignment (Figure 1). Briefly, among those 37 proteins, 15 are most likely active DGCs, 18 could be active PDEAs, 1 protein (CD0522) could either be an active DGC and/or PDEA, and 3 contain a predicted catalytically inactive GGDEF domain (Table S1). Except for CD0522, all the proteins with both GGDEF and EAL domains were predicted to be PDEAs having an inactive GGDEF partner domain since these have a degenerated GG[D/E]EF motif. Most of the 37 proteins are predicted to have at least one sensor domain, transmembrane regions or a signal peptide region.

Distribution of the putative c-di-GMP–signalling components in *C. difficile*

To assess whether c-di-GMP signalling components are conserved within the species *C. difficile*, orthologs of the proteins identified in strain 630 were exhaustively sought for in 19 other partially or completely sequenced *C. difficile* genomes using tblastn. Most genes found in *C. difficile* 630 are conserved in all the strains examined with 31 of the 37 GGDEF and/or EAL proteins having an ortholog in at least 15 strains and only CD2753 and CD2754 being unique to strain 630 (Figure 2A). The putative glycosyl-transferase CD2545, the sole protein predicted to contain a PilZ domain (Pfam PF07238), has orthologs in all the other analyzed strains (Figure 2A). Furthermore, two RNA targets (named herein Cd1-a and 84Cd, respectively) that have been shown to bind c-di-GMP, the riboswitch Cd1 located upstream of the flagellum synthesis operon, and the self-splicing group I intron (tandem riboswitch-ribozyme) are conserved in a subset of strains [22,23].

Active site
guanyl binding
Active site

I_S   Active site   I_P   Active site   I_P

313 316 327 335 344 359 368 390

PleD	REC REC	R D N D RxxD GGEEF R	454
CD0537	REC	D N D RxxD GGEEF R	319
CD0583		D N D GGDEF	409
CD0584		D N D GGDEF	392
CD0707	SBP Bac 3 SBP Bac 3	D N D GGDEF	837
CD0997	PAS 3 PAS 3 PAS 3	D N D RxxD GGDEF	1184
CD1015	PAS 3	D N D GGDEF	537
CD1028		R RxxD QKDMI R	282
CD1185		D N D GGEEF	572
CD1419	PAS 3 PAS 3	D N N GGDEF	577
CD1420		D N D GGDEF	281
CD1841	PAS 3 PAS 3 PAS 3	D N D RxxD GGDEF	1145
CD2244		D N D GGEEF	391
CD2384		D N D GGEEI R	632
CD2385		D N D GGEEF	613
CD2754		D N D GGDEF	516
CD2887		D N D GGEEY	384
CD3365		D N D GGDEF	318
CD3625		D N D RxxD GGDEF	835

c-di-GMP binding
metal binding
c-di-GMP binding
metal binding
active site
general base

161 175 233 265 295 316 352

RocR	REC	Q EVLxR N E D K E	392
CD0204		N N N DADLF Q ESLxR N E D K E	744
CD0522		R D N D YADVF Q EALxR N E D K E	1032
		D N D GGDEF	
CD0748		D N D SGEEF Q EALxR N E D K E	1011
CD0757	PAS 3	D N D DGDEF Q EVLxR N E D K E	556
CD0811		A N D GGDIF Q EALxR N E D K E	642
CD1421		N N E KGEGF Q EALxR N E D K E PAS 3	744
CD1476	SBP Bac 3 SBP Bac 3	D N N SADHF R Q EALxR N E D K E	965
CD1515	PAS 3	D N D DGDEM Q EVLxR N E D K E	697
CD1538		D N N DNDHF R Q EALxR N E D K E	739
CD1616		D N N SADNF R Q EALxR N E D K E	514
CD1651		D N D IADRF R Q EALxR N E D K E	619
CD1840		D N N SADNF Q EALxR N E D K E	502
CD2134		D N D GSDNF Q EALxR N E D K E	615
CD2663	cache1	D N D SNDNF Q EALxR N E D K E	759
CD2753		D N D NGDVF Q EALxR N E D K E	900
		R SEEGF	
CD2760		D N D KSDVF Q EALxR N E D K E	515
CD2873		N N D SNDQF Q EALxR N E D K E	759
CD2965	cache1	D N D SSDYF Q EALxR N E D K E	739
CD3650	PTS EIIC	Q EGLxR N E D K E	700

Figure 1. Domain composition and organization of the 37 putative c-di-GMP–signalling proteins encoded by *C. difficile* 630. Proteins names are listed on the left and predicted sizes in amino acids are indicated on the right. GGDEF domains are shown in green and EAL domains are shown in red. The DGC PleD from *C. crescentus* (YP_002517919) and the PDEA RocR from *P. aeruginosa* (NP_252636) are shown as references to highlight amino acids important for enzymatic activity. The putative functions and positions of important amino acid residues [28,29] are listed above each domain in black (conserved) or grey (non-conserved) and their position in PleD and RocR are indicated. Blue boxes are predicted signal peptide regions. Black boxes represent predicted transmembrane regions. Grey boxes represent predicted coiled-coil motifs. Additional sensor domains predicted by Pfam 24.0 are also shown in white. Proteins are not drawn to scale. Cache 1, Calcium channels and chemotaxis receptor family 1; PAS 3, PAS fold family 3; PTS EIIC, Phosphotransferase system EIIC; REC, Response regulator receiver; SBP bac 3, Bacterial extracellular solute-binding proteins, family 3; I_P, primary inhibitory site; I_S, secondary inhibitory site.
doi:10.1371/journal.pgen.1002039.g001

The conservation pattern of the c-di-GMP regulatory proteins and downstream effectors clearly follows the phylogenetic distribution of the strains (Figure 2A and 2B). Interestingly, the cluster of 10 hypervirulent NAP1/BI/027 strains (cluster CD196/2007855 in Figure 2B) regroups the strains encoding the lowest number of c-di-GMP signalling components.

Like CD2753 and CD2754 that are specific to *C. difficile* 630, we assumed that other strains could also encode c-di-GMP signalling proteins that were absent from all of the other strains. The Microbial Signal Transduction database (MiST$_2$) [30], which currently provides data for 14 out of our 19 strains, contains 3 putative c-di-GMP turn-over proteins that are absent from *C. difficile* 630's proteome. Notably, CdifA_020200002673, is unique to strain ATCC43255 (Figure 2A), and is the only HD-GYP domain-containing protein detected in *C. difficile*. Based on analysis of amino acid conservation, it is predicted to have both DGC and PDE enzymatic activities (Figure S1). To identify additional strain-specific genes encoding c-di-GMP signalling proteins, we performed profile hidden-Markov model (profile HMM) searches with the Pfam HMMs for GGDEF, EAL, HD and PilZ conserved domains using HMMER3 software against the proteomes of strains 2007855, BI1, BI9, CF5, M120 and M68 as predicted by GeneMark.hmm (Table S2). No additional c-di-GMP-signalling proteins were detected. Besides CD2545, no other PilZ domain-containing protein was found. Finally, no GGDEF, EAL, HD-GYP or PilZ domain-containing proteins were found to be encoded by *C. difficile* plasmids pCD630, pCD6, pCDBI1 or the 300 kb putative phage or plasmid from BI1 (Table S2).

The remarkable prevalence of c-di-GMP-signalling components among *C. difficile* strains suggests that c-di-GMP is an important second messenger in this bacterium. The functionality of the 31 most conserved GGDEF and/or EAL domain proteins among *C. difficile* strains (Figure 2A) was therefore assessed to confirm their biological activity.
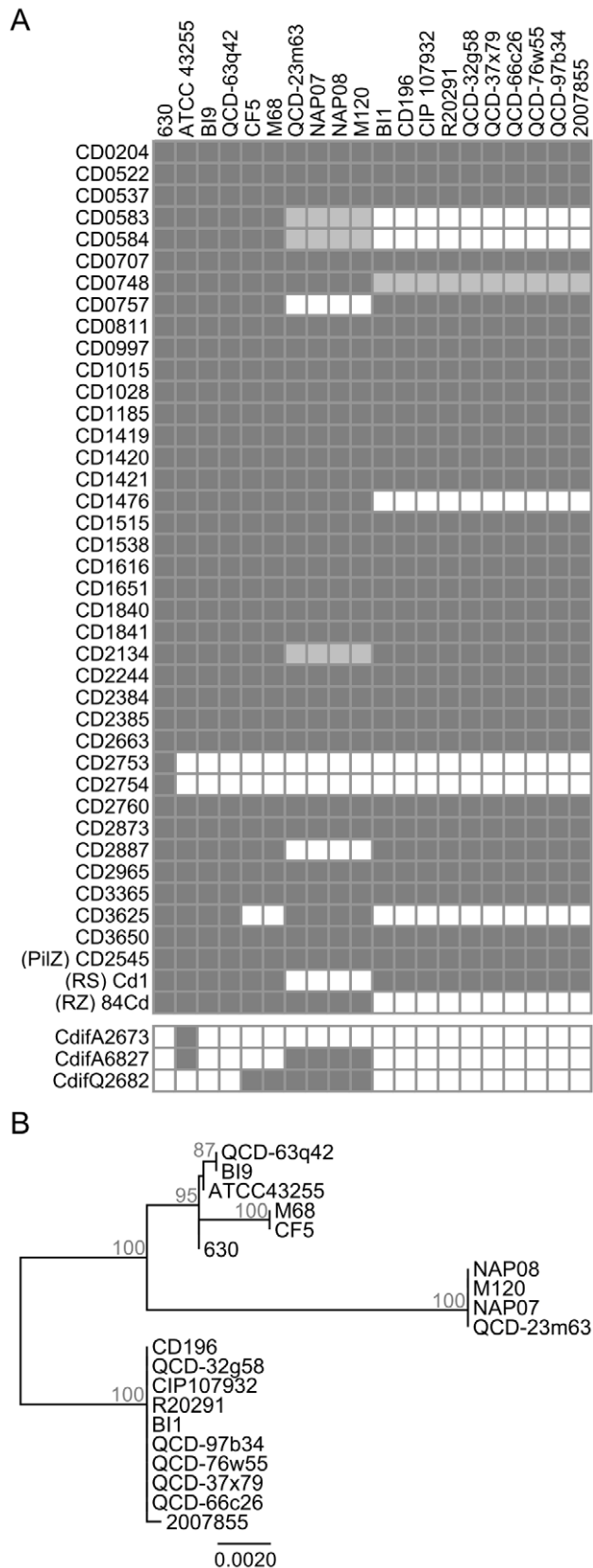
Heterologous expression of DGC and PDEA encoding genes in *V. cholerae*

To the best of our knowledge, no model of Gram-positive bacteria is currently available to efficiently and reliably evaluate *in vivo* the enzymatic activity of proteins regulating the intracellular levels of c-di-GMP. Moreover, to circumvent the tedious laboratory procedures associated with working with *C. difficile* and the lack of information regarding the phenotypes regulated by c-di-GMP in this bacterium, assessment of the enzymatic activity of the 31 putative DGCs and PDEAs (cloned from *C. difficile* 630) was carried out by heterologous expression in *V. cholerae*. Characteristic phenotype alterations associated with variations of the intracellular c-di-GMP concentration of *V. cholerae* are easily observable and measurable. High levels of intracellular c-di-GMP concentrations increase biofilm formation and decrease motility whereas lower c-di-GMP concentrations cause the opposite effects [31]. As expected, ectopic expression of most of the predicted DGCs containing the canonical GG[D/E]EF motif decreased cell motility (Figure 3A) and increased biofilm formation (Figure 3C),

supporting the hypothesis that these *C. difficile* proteins are genuine and functional DGCs (Table S1). Replacement of the second glycine of the GGDEF motif by site-directed mutagenesis has been shown to greatly reduce the enzymatic activity of other DGCs [32,33]. Such a mutation in CD1420 (mutant G204E) abolished alterations of both biofilm and motility phenotypes (Figure 4A, 4B and 4C) confirming that the phenotypical alterations observed in *V. cholerae* are linked to the enzymatic activity of this functional *C. difficile* DGC. Moreover, the expression of CD1420 in *V. cholerae* N16961 led to a dramatic increase of the c-di-GMP level compared to the same strain expressing LacZ (Figure S2B and S2D, Text S1). Furthermore, we observed that CD1015, CD1185 and CD1420, the DGCs causing the strongest phenotypical shifts (Figure 3A and 3C), lacked the amino acid motif RXXD that is part of the retro-inhibition site (I-site) of the GGDEF domain (Figure 1) [34]. The putative DGC CD2887 significantly enhanced biofilm formation without affecting the motility of *V. cholerae*. Interestingly, CD2887 contains a GGEEY motif instead of the canonical GG[D/E]EF motif suggesting that it might not be a functional DGC. However, Malone and colleagues [35] showed that upon substitution of the F amino acid residue for a Y residue in the A-site, the DGC WspR of *Pseudomonas fluorescens* retains its activity. Unlike the putative DGCs presented above, several predicted DGCs that appear to possess all the conserved amino acid residues necessary for c-di-GMP synthesis (e.g. CD2385 and CD3365) did not modulate biofilm formation or motility of *V. cholerae* in our experimental conditions. Yet we cannot conclude that these putative DGCs are not functional since they might not have been produced, may have been unstable or may lack the appropriate activating signal in the heterologous host. Unexpectedly, CD0537, which also has a canonical A-site, did not enhance biofilm formation but enhanced cell motility by ~60%. While this increase is notable compared to other DGCs, it remained very modest compared to the increase promoted by PDEAs (see below) suggesting that in our experimental setting, this putative DGC was not functional. The unexpected result on motility could result from partial sequestration of intracellular c-di-GMP by CD0537's I-site upon overexpression of the protein. Additionally, the lack of apparent DGC activity of CD0537's might simply be due to a lack of phosphorylation of its phosphoreceptor REC domain, a modification that could be catalyzed by the putative kinase *cheA* (CD0539) that was absent in our assay. Other DGCs have been shown to be activated by phosporylation of their phosphoreceptor REC domain [29,32,36]. As expected, CD1028 and CD2384, which respectively possess the degenerated QKDMI and GGEEI motifs, did not alter *V. cholerae*'s phenotypes. Consistent with our observation, the substitution of the F residue of the A-site for an I residue alone is known to eliminate the activity of WspR of *P. fluorescens* [35].

Expression of several putative PDEAs (CD0757, CD1515, CD1616, CD1840 and CD2663) significantly enhanced cell motility on soft agar by 2 to 4 fold (Figure 3B and Table S1). While five putative PDEAs exhibited significant activity (CD0204, CD1421, CD1515, CD2134 and CD2663), in most cases, the

A

Column headers: 630, ATCC 43255, BI9, QCD-63q42, CF5, M68, QCD-23m63, NAP07, NAP08, M120, BI1, CD196, CIP 107932, R20291, QCD-32g58, QCD-37x79, QCD-66c26, QCD-76w55, QCD-97b34, 2007855

Row labels:
CD0204
CD0522
CD0537
CD0583
CD0584
CD0707
CD0748
CD0757
CD0811
CD0997
CD1015
CD1028
CD1185
CD1419
CD1420
CD1421
CD1476
CD1515
CD1538
CD1616
CD1651
CD1840
CD1841
CD2134
CD2244
CD2384
CD2385
CD2663
CD2753
CD2754
CD2760
CD2873
CD2887
CD2965
CD3365
CD3625
CD3650
(PilZ) CD2545
(RS) Cd1
(RZ) 84Cd

CdifA2673
CdifA6827
CdifQ2682

B

Phylogenetic tree:
87 — QCD-63q42
BI9
95 — ATCC43255
100 — M68
CF5
100 — 630
100 — NAP08
M120
NAP07
QCD-23m63
CD196
QCD-32g58
CIP107932
R20291
BI1
100 — QCD-97b34
QCD-76w55
QCD-37x79
QCD-66c26
2007855

0.0020

Figure 2. Conservation of c-di-GMP–signalling components of *C. difficile* 630 in other strains of *C. difficile*. (A) Protein orthologs exhibiting ≥90% identity are indicated by dark grey boxes. More divergent proteins sharing between 85 and 89% identity or containing major gaps are indicated by light grey boxes. Proteins that are absent are indicated by empty boxes. The c-di-GMP-sensing riboswitch Cd1 ((RS) Cd1) and 84Cd aptamer from the tandem riboswitch-ribozyme ((RZ) 84Cd) are indicated by a dark gray box when they shared ≥90% and were in the same genetic context as in strain 630. CdifQ2682, CdifA2673, and CdifA6827 were used to refer to proteins CdifQCD-2_020200002682 (ZP_05400016) from *C. difficile* QCD-23m63, CdifA_020200002673 (ZP_05349638) and CdifA_020200006827 (ZP_05350459) from *C. difficile* ATCC 43255, respectively. (B) Phylogenetic tree of *C. difficile* isolates based on *rpoB* nucleotide sequences. Bootstrap values are indicated at branch point in percent. Only values ≥70% are shown.
doi:10.1371/journal.pgen.1002039.g002

impact of ectopic expression of these proteins on biofilm formation was modest and not significant (Figure 3D). The weak response of *V. cholerae* N16961 to PDEA activity could be due to its low basal level of biofilm formation in our assays. However, we indirectly confirmed that the *V. cholerae* motility response to overexpression of these proteins was linked to PDEA activity by mutating the glutamic acid residue of the EVLxR motif of CD0757, which is critical for enzymatic activity. Indeed, unlike the wild-type protein, overexpression of CD0757-E339A did not enhance the motility of *V. cholerae* N16961 on soft agar (Figure 4A and 4C).

CD0522 has dual enzymatic activity

Since CD0522 has a particular combination of GGDEF and EAL domains suggesting that it could have both DGC and PDEA activities, it was analyzed apart from those in Figure 3. CD0522 contains two predicted N-terminal GGDEF domains and one predicted C-terminal EAL domain (Figure 1 and Figure 4D). Our analysis indicates that the first GGDEF domain and the EAL domain should be catalytically active due to the conservation of the A-sites. However, the second GGDEF domain contains the strongly degenerated motif YADVF suggesting that it is catalytically inactive (Figure 1 and Figure 4D). CD0522 or its individual domains were tested in our *V. cholerae* heterologous expression model to verify these predictions (Figure 4C, 4E and 4F). Since the variations of cell motility and biofilm formation we observed upon ectopic expression of CD0522 were not statistically significant, we could not clearly establish a DGC activity for the complete CD0522. On the other hand, expression of the N-terminal fragment N-448, which encompasses the first GGDEF domain, reduced motility by more than half and increased biofilm by ~7 fold, suggesting that this fragment of CD0522 acts as a functional DGC (Figure 4E and 4F). The observed phenotypes correlate with a marked increase of intracellular c-di-GMP upon expression of N-448 in *V. cholerae* N16961 compared to the same strain expressing LacZ (Figure S2B and S2C, Text S1). DGC catalytic activity of CD0522 was also indirectly confirmed by mutating the glutamic acid residue of the EAL motif to abolish any possible PDEA activity (Figure 4D). Expression of CD0522-E814A altered phenotypes as expected for a DGC, as observed for N-448 (Figure 4C, 4E and 4F). Expression of the C-terminal fragment C-307 led to a modest but significant increase of motility. As expected, expression of the degenerated GGDEF-containing central fragment M-305 did not lead to any significant change of phenotype compared to the control. The degenerated GGDEF domain of M-305 could be involved in the regulation of the PDEA activity of the EAL domain of CD0522. We observed that motility of cells expressing MC-591, which contains the central and C-terminal fragments, increased by ~50% (Figure 4E), which is consistent with a diminution of intracellular c-di-GMP. Increased motility was also observed when we overexpressed CD0522-G366E, a protein containing a substitution of the second glycine
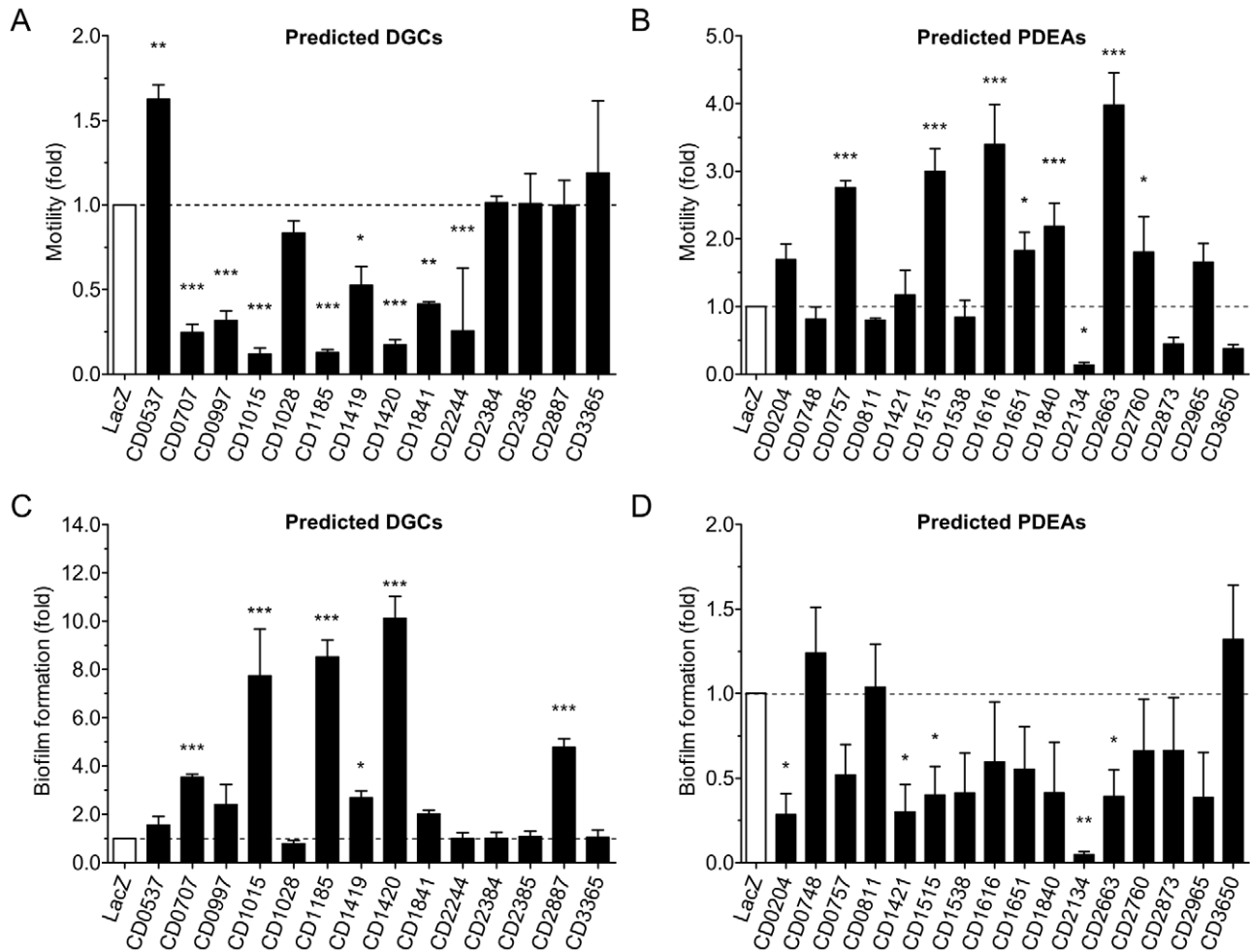
Figure 3. Phenotypic changes triggered by the expression of *C. difficile* 630 genes encoding predicted DGCs or PDEAs in *V. cholerae* N16961. Relative motility of cells over-expressing (A) predicted DGCs or (B) predicted PDEAs. Relative biofilm formation of mutant cells over-expressing (C) predicted DGCs or (D) predicted PDEAs. *V. cholerae* over-expressing *lacZ* was used as a reference control. The means and standard deviations obtained from at least three independent assays are shown. One-way ANOVA with a Dunnett multiple comparison post-test was used to compare the strains over-expressing putative c-di-GMP signalling proteins and the control strain (LacZ) (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).
doi:10.1371/journal.pgen.1002039.g003

residue of the first GGDEF to abolish any DGC activity (Figure 4C and 4E). Complex proteins such as CD0522 that are composed of several GGDEF and EAL domains suggest a possible two-way c-di-GMP control and must be studied in detail to reveal what stimuli switches their enzymatic activity between the DGC or PDEA state.

In vitro enzymatic assays

CD1420 and CD0757 enzymatic activities were further assessed *in vitro* to corroborate the results obtained in the *V. cholerae* expression model and confirm that the *C. difficile* proteins are genuine DGC and PDEA, respectively. These proteins were chosen for their strong activity in *V. cholerae* and the simplicity of the structure of the N-terminal sensor domain that suggested little requirements for *in vitro* assays (Figure 1). Purified CD1420 in its native form was able to produce c-di-GMP from GTP as substrate and the accumulation of the product increased with time (Figure 5A). Conversely, purified CD0757 did not produce c-di-GMP from GTP even after 1 h incubation. Therefore, we confirmed that CD1420 is a functional DGC. The absence of DGC activity of CD0757 suggests that it contains an inactive GGDEF domain and acts as a PDEA only.

CD0757 was then assessed for PDEA activity on c-di-GMP. c-di-GMP hydrolysis by PDEAs is known to yield the linear diguanylate pGpG [37]. We incubated purified CD0757 with radiolabeled c-di-GMP, yielding small amounts of pGpG. This characteristic PDEA activity was abolished by denaturing the protein prior to the assay (Figure 5B). Inactive GGDEF domains have been shown to enhance PDEA activity of an adjacent EAL domain by binding GTP [26]. Addition of GTP to the enzymatic reaction increased noticeably the PDEA activity of CD0757 presumably through binding to the GGDEF domain like for PdeA (CC3396) from *C. crescentus*. After a 30-min incubation period, virtually all the c-di-GMP was converted to pGpG. Marginal degradation of c-di-GMP to GMP by CD0757 was detected as previously shown to occur with another PDEA [37].

Discussion

Studies on c-di-GMP have addressed with some depth many aspects regarding the proteins involved in its synthesis (DGCs and PDEAs) and the molecular targets of c-di-GMP such as proteins and riboswitches in several bacteria. While c-di-GMP signalling has been extensively studied in many Gram-negative bacteria like
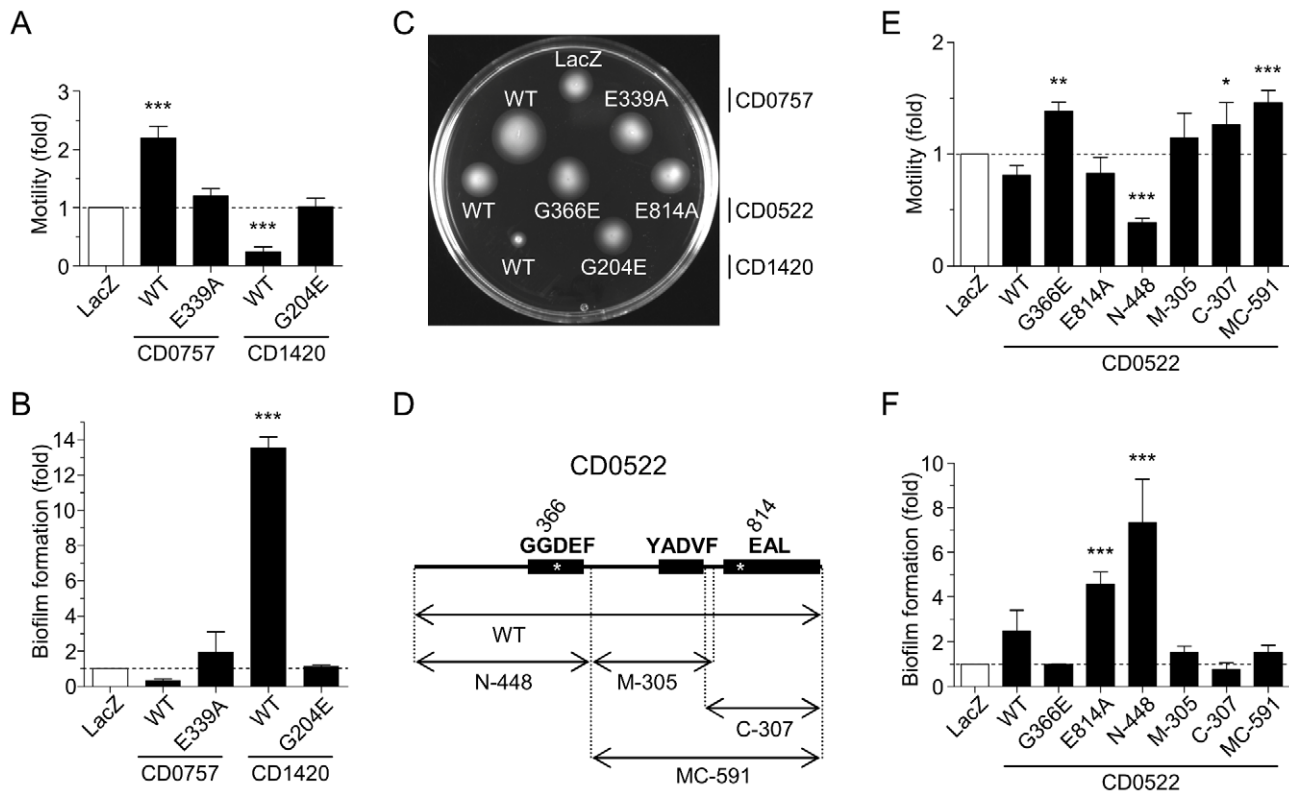
Figure 4. Alteration of *V. cholerae* N16961 motility and biofilm formation upon expression of wild-type or mutated putative DGCs and PDEAs from *C. difficile* 630. (A) Relative motility upon expression of wild-type or mutated CD0757 or CD1420. (B) Relative biofilm expression upon expression of wild-type or mutated CD0757 or CD1420. (C) Motility of the strain expressing the wild-type (WT) and mutated proteins on LB soft agar plates at 30°C. The picture shown is representative of four independent experiments. (D) Schematic representation of CD0522 protein fragments expressed. The amino acids in the signature motifs of the GGDEF or EAL domains are indicated above each domain. Positions of the mutations in the GGDEF or EAL domains are indicated by asterisks. (E) Relative motility upon expression of full-length, mutated or truncated CD0522. (F) Relative biofilm formation upon expression of full-length, mutated or truncated CD0522. The means and standard deviations obtained from four independent assays are shown. One-way ANOVA with a Dunnett multiple comparison post-test was used to compare the cells over-expressing putative c-di-GMP signalling proteins and the control strain (LacZ) (*, P<0.05; **, P<0.01; ***, P<0.001).
doi:10.1371/journal.pgen.1002039.g004

C. crescentus, *E. coli*, *V. cholerae*, *Salmonella* and *Pseudomonas*, very few studies have been carried out on Gram-positive bacteria. To the best of our knowledge, the staphylococcal GGDEF domain protein GdpS has been the only c-di-GMP regulatory protein studied to date in low G+C Gram-positive bacteria. GdpS does not seem to have any measurable DGC activity [16]. The recent discovery of a functional c-di-GMP binding riboswitch in *C. difficile* and *Bacillus cereus*, as well as the prediction of several other similar riboswitches in other Gram-positive bacteria has revived the interest in studying c-di-GMP metabolism in these microorganisms. The recent characterization of a c-di-GMP-dependent self-splicing group I ribozyme in *C. difficile* further reinforces the role of c-di-GMP in Gram-positive bacteria.

In this work, we have shown that many of the genes encoding putative DGCs and PDEAs of *C. difficile* behave like genuine DGCs and PDEAs in heterologous expression experiments (Figure 3 and Figure 4, Table S1). This number of c-di-GMP regulatory proteins encoded by *C. difficile* is high compared to what is found in its closest relatives (Table S3), and also among the Firmicutes in general (median = 1) [38]. Analysis of the genomes of 49 strains of Clostridiaceae representing 27 species revealed that most contain less than 20 of such genes (Table S3). Only 2 species of *Clostridium* were found to encode more putative DGCs/PDEAs than *C. difficile*, *Clostridium asparagiforme* DSM15981 and *Clostridium bolteae* ATCC BAA-613, two newly characterized yet barely

studied species isolated from human fecal samples [39,40]. The disparity in the occurrence of c-di-GMP signalling proteins is remarkable. The two species coding for the lowest number of c-di-GMP regulatory proteins, *Clostridium hiranonis* and *Clostridium bartlettii*, are the closest phylogenetically related species to *C. difficile* (Figure S3 and Table S3). On the opposite, the two species coding for the highest number of GGDEF, EAL or HD-GYP protein, *C. asparagiforme* and *C. bolteae*, are among the most distant species from *C. difficile*. Additionally, *C. difficile*, which encodes with one exception no HD-GYP domain proteins, seems to be an exception among the Clostridiaceae and contrasts with *Clostridium beijerinckii* which encodes 14 HD-GYP domain proteins (Table S3), while retaining the same number of c-di-GMP regulatory proteins as *C. difficile*. In addition, *C. difficile* does not seem to carry any gene encoding c-di-GMP-signalling proteins that could have been recently exchanged by horizontal transfer with the 3 other *Clostridium* species containing the highest number of such proteins (Table S3 and data not shown). The Clostridiaceae seem to have a high number of c-di-GMP-signalling proteins among the Firmicutes in general, but it remains similar to other Firmicutes of comparable size (3000–4000 genes, median = 10) [38].

The high number of c-di-GMP turn-over proteins in *C. difficile* is likely indicative of the importance of this second messenger in the bacterium's lifecycle and suggests a major role in regulating different phenotypes. The diversity of N-terminal structures
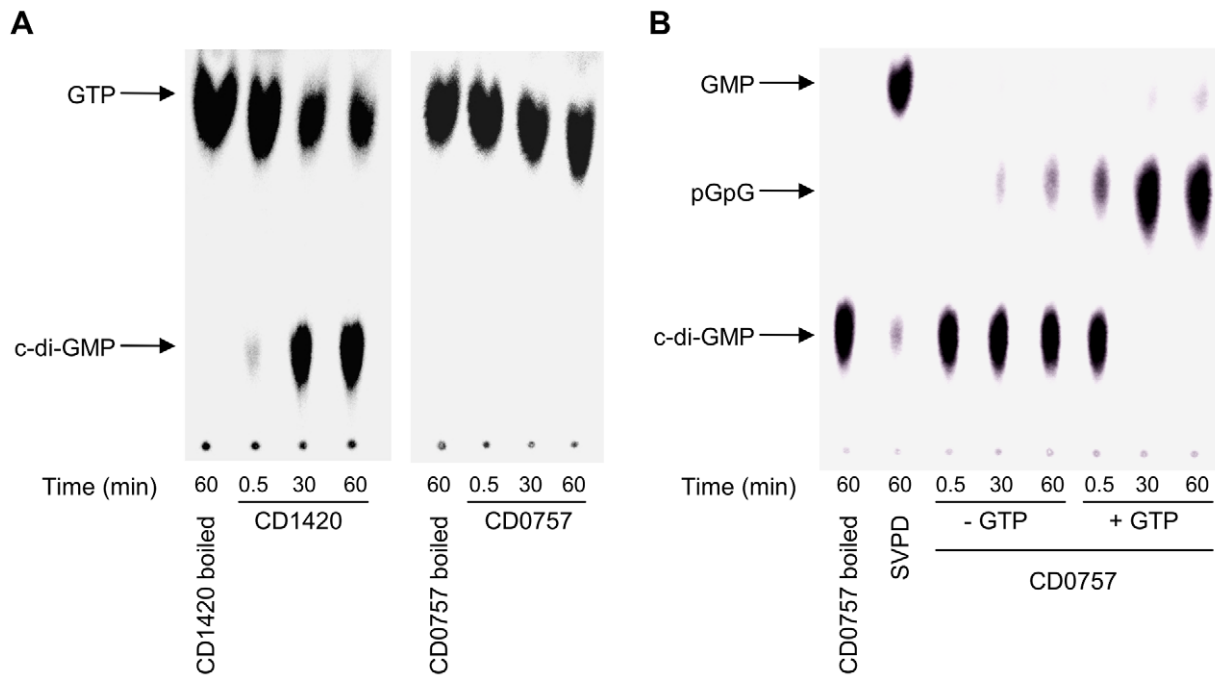
Figure 5. TLC analysis of the enzymatic activities of CD1420 and CD0757. (A) DGC activity of purified CD1420. (B) PDEA activity of purified CD0757. Snake venom phosphodiesterase (SVPD) was used as a positive control.
doi:10.1371/journal.pgen.1002039.g005

suggests that their function is not redundant. Instead, these proteins could individually act in a functionally or spatially sequestered way, in addition to being temporally regulated through differential expression. It has been shown that DGCs usually are not interchangeable and can contribute to very specific and distinct phenotypes for a unique microorganism. For example, while the DGC YddV of *E. coli* impacts poly-N-acetylglucosamine production, other DGCs like AdrA do not [41]. Instead AdrA controls the production of cellulose, another exopolysaccharide, in *E. coli* and *Salmonella* [42,43]. Furthermore, the prevalence of DGCs and PDEAs in *C. difficile* could also indicate the importance of these proteins in sensing and relaying a diversified array of environmental conditions through their sensor domains or their eventual differential expression. In *V. cholerae*, the PDEA CdpA is not expressed *in vivo* until the late stage of infection in a mouse colonization model [44]. *C. difficile* vegetative cells encounter various environmental conditions during their journey through the gastrointestinal tract, during which c-di-GMP signalling might play a role in regulating diverse phenotypes.

Despite the current lack of experimental data, it is reasonable to assume that c-di-GMP regulates at least two phenotypes in *C. difficile*: flagella synthesis/motility and polysaccharide synthesis. A putative c-di-GMP-binding PilZ domain is located in the putative glycosyltransferase CD2545, which is predicted to be a cellulose synthase. Interestingly, although motility is commonly controlled by c-di-GMP in bacteria, the c-di-GMP-responsive effectors that likely regulate this phenotype in *C. difficile* appear to differ from those found in *V. cholerae*, *E. coli* and related bacteria. The c-di-GMP-sensing riboswitch Cd1 appears to control the transcription of the large operon of genes essential for assembling the flagellum apparatus [22]. Bacterial flagella are obviously important for motility but can also be involved in adhesion. Adhesion to mouse cecal mucus of the flagellin FliC and of the flagellum cap protein FliD of *C. difficile* has been demonstrated *in vitro* [45]. Moreover, FliD has been shown to specifically adhere to cultured cells [45].

Therefore, c-di-GMP signalling might impact both cell motility and adhesion of *C. difficile* to mucosal surface through the regulation of flagellum assembly. Additionally, c-di-GMP signalling may play a significant role in the excessive inflammation caused by *C. difficile* infection since flagellin is a very potent immunogenic protein recognized as a proinflammatory ligand by toll-like receptor 5 (TLR-5) located at the baso-lateral surface of intestinal cells (reviewed in [46]).

Except for one EAL protein (CD3650), all of *C. difficile*'s PDEAs contain a GGDEF domain predicted to be non-catalytically active as shown *in vitro* for CD0757 (Figure 1 and Figure 5). Composite proteins containing both GGDEF and EAL domains are relatively frequent, representing approximately one third of proteins with such domains [47]. Some of these composite proteins act as DGCs, PDEAs, or have both activities (reviewed in [47]). Inactive GGDEF or EAL domains can act as sensor domains rather than catalytic domains. GGDEF domain proteins with degenerated active sites have been reported to bind c-di-GMP at their conserved I-site, as for the *C. crescentus* protein PopA [48], or to retain the ability to bind GTP at their degenerated active site, as for the *C. crescentus* protein PdeA [26]. Christen and colleagues [26] have demonstrated that binding of GTP to the inactive GGDEF domain of PdeA of *C. crescentus*, strongly enhanced the PDEA activity of the C-terminal EAL domain. The authors formulated two hypotheses to explain why the PDEA activity is linked to GTP intracellular concentrations: (i) to prevent GTP pools to drop by the uncontrolled successive activities of DGCs and PDEAs and (ii) to sense physiological changes. Intracellular GTP levels have been shown to impact the activity of CodY, a major transcriptional regulator in many low G+C Gram-positive bacteria such as *C. difficile*, *S. aureus*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus mutans*, *Listeria monocytogenes*, *B. cereus* and *Bacillus anthracis* in which it affects virulence gene expression ([49] and references therein). CodY is known to have greater affinity to target DNA promoter regions upon binding of two synergistic

effectors, GTP and branched-chain amino acids [50,51]. *C. difficile* CodY has been shown to repress the expression of the toxin A and B genes (*tcdA* and *tcdB*), through binding to the promoter region of the positive transcriptional regulator TcdR [52]. A recent study aimed at identifying all DNA promoter regions targeted by *C. difficile* CodY as well as genes differentially expressed in a *codY* null mutant [49]. Interestingly, among the 165 genes identified with altered expression, PDEA genes CD0757 and CD1476 were highly derepressed. Additionally, DNA regions containing CD1476, CD2385, CD2873, CD2965 and CD3650 were identified as CodY binding-sites. These data suggest a probable interplay between the c-di-GMP and CodY signalling pathways, known to be important in the regulation of many metabolic genes and of the major virulence factors, toxins A and B [49,52].

To the best of our knowledge, no model of Gram-positive bacteria is currently available to efficiently and reliably evaluate the enzymatic activity of proteins regulating the intracellular levels of c-di-GMP. With the recent availability of molecular tools for *C. difficile* genetic manipulation [53–55], it will finally be possible to study in detail the many genes involved in c-di-GMP signalling and turn-over in this bacterium and to identify the phenotypes associated with the variation of intracellular c-di-GMP pools. The need to decipher the regulatory mechanisms underlying *C. difficile*'s behaviors is imperative to the development of new therapeutics and treatment strategies. Particularly, the bacterial signalling pathways and phenotypes involved at the colon mucosal interface ought to be addressed.

Materials and Methods

Bioinformatics

Proteins containing the c-di-GMP-associated conserved domains (GGDEF, EAL, HD for HD-GYP and PilZ) were searched for in the Clostridiaceae proteomes on the Pfam 24.0 server [25]. HD domains were further analyzed to identify HD-GYP domains by looking for the HD-GYP amino acid motif by multiple alignment with the HD-GYP domain of Rpfg from *Xanthomonas campestris* 8004 (Accession number AAY49388) using ClustalW version 2.0.12 [56]. Other conserved domains, signal peptides, transmembrane regions, and coiled-coil motifs annotations are as determined on Pfam 24.0 [25]. Identification of proteins with c-di-GMP-associated conserved domains in *C. difficile* strains other than 630 was achieved with the hmmsearch program of the HMMER 3.0 software (http://hmmer. org/). *C. difficile* annotated protein sequences were retrieved for the 13 other strains and 2 plasmids available in the NCBI Refseq database (Table S2) [57]. Protein sequences from *C. difficile* 2007855, BI1, BI9, CF5, M120 and M68 genomes and extrachromosomal sequences (Table S2) were predicted using Gene-Mark.hmm for Prokaryotes version 2.4 [58]. Profile hidden Markov models (profile-HMMs) of c-di-GMP-associated conserved domains were downloaded from Pfam 24.0. The bit score threshold values used in every search were the "trusted cutoff" values for the Pfam profile-HMMs. Proteins containing the c-di-GMP-related conserved domains identified using HMMER 3.0 software were further analyzed to identify other conserved domains (Pfam 24.0), signal peptides and transmembrane regions (Phobius [59]), and coiled-coil motifs (ncoils [60]). Nucleotide and amino acid conservation of selected *C. difficile* 630 genes and proteins were assessed with the appropriate BLAST algorithms [61]. Since most of the genomes are drafts, pseudogenes were ignored and assumed to be the results of sequencing errors. As a matter of fact, pseudogenes are found in many of these strains even for important, unique and well-conserved genes such as the gene encoding DNA polymerase I (data not shown).

Phylogenetic trees were generated using the neighbor-joining method as implemented by ClustalX version 2.012 [56] from gapless alignments of nucleotide sequences. Nucleotide sequences were aligned using ClustalW version 2.0.12 [56] and gap columns were removed using Jalview version 2.5 multiple alignment editor [62]. The reliability of each tree was subjected to a bootstrap test with 1000 replications. Trees were edited using FigTree version 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Growth conditions

Bacterial strains were routinely grown in Luria–Bertani (LB) broth at 37°C in an orbital shaker and maintained at −80°C in LB broth containing 15% (v/v) glycerol. Ampicillin (Ap) was used at 100 μg ml^{-1} when needed. For induction of gene expression in the strains carrying arabinose-inducible vectors (pBAD series), L-arabinose was added to the growth medium at a final concentration of 0.02% (w/v).

Bacterial strains and plasmid construction

The bacterial strains and plasmids used in this study are described in Table S4. The oligonucleotides used for plasmid constructions are described in Table S5. For expression of putative DGCs and PDEAs in *V. cholerae*, genes cloned in pBAD-TOPO were amplified by PCR with their native Shine-Dalgarno sequence using *C. difficile* 630 genomic DNA as a template. Truncated versions of CD0522 were cloned to include the native Shine-Dalgarno sequence of CD0522. DNA was amplified to express CD0522 protein fragments N-448, M-305, C-307 and MC-591 respectively containing the N-terminal 448 amino acids (aa), 305 aa encompassing the middle domain, the 307 aa in C-terminal and 591 aa encompassing the middle and C-terminal domains (Figure 4D).

Plasmids pCD0522-G366E, pCD0522-E814A, pCD0757-E339A and pCD1420-G204E, which accordingly contain amino acid substitutions in their respective conserved GG[D/E]EF or EXL A-sites, were created by site-directed mutagenesis of pCD0522, pCD0757 and pCD1420 using the QuickChange Lightning Site-directed Mutagenesis Kit (Stratagene) using primer pairs listed in Table S5. The mutations introduced were designed to create new EarI or PvuII restriction sites for initial screening of the mutated plasmids. Mutated genes were verified by sequencing.

For CD0757 and CD1420 proteins purification, the corresponding genes were amplified by PCR from *C. difficile* 630 genomic DNA and cloned into BamHI/SalI-digested pGEX6P-1 in frame with the glutathione S-transferase (GST) coding sequence.

Molecular biology methods

All the enzymes used in this study were obtained from New England BioLabs and were used according to the manufacturer's instructions. Plasmid DNA was prepared with a Qiaprep Spin miniprep kit (Qiagen). Genomic DNA of *C. difficile* 630 was extracted using the illustra bacteria GenomicPrep mini spin kit (GE Healthcare). PCR assays were performed with the primers described in Table S4 in 50 μl of PCR mixtures with 1 U of *Pfu* Ultra DNA polymerase (Agilent). PCR conditions were as follows: (i) 3 min at 94°C, (ii) 30 cycles of 30 s at 94°C, 30 s at suitable annealing temperature, and 30–300 s at 72°C, and (iii) 5 min at 72°C. When needed PCR products were purified using a QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions. *E. coli* was transformed by electroporation according to Dower and colleagues [63]. *V. cholerae* was transformed by electroporation according to Occhino and colleagues [64]. In both cases, transformation was carried out in

0.1 cm electroporation cuvettes using a Bio-Rad GenePulser Xcell apparatus set at 25 μF, 200 Ω and 1.8 kV.

Motility and biofilm formation assays

Motility and biofilm assays were performed as described before [32]. Briefly, a semi-solid medium composed of 1% tryptone, 0.5% NaCl, 0.3% agar supplemented with ampicillin and L-arabinose was used to evaluate motility of *V. cholerae* mutant strains during over-expression assays at 30°C. Motility was assessed from the comparison of the surface area (mm²) of the colonies from plate images captured and analyzed using a Gel Doc XR system and Quantity One software (Bio-Rad). The capacity of *V. cholerae* mutant strains to form biofilm was determined after 6 h static growth in LB broth containing ampicillin and L-arabinose at 30°C. Bound crystal violet was solubilized with 200 μl of 95% ethanol and quantified by absorbance at 595 nm in a Model 680 microplate reader (Bio-Rad). Motility and biofilm formation assays were carried in triplicate and data were normalized as fold expression compared with the control LacZ over-expressing bacteria. Data from at least three independent experiments were combined.

Production of recombinant proteins

Overnight-grown cultures of *E. coli* BL21 bearing pGCD0757 or pGCD1420 were diluted 1:100 in fresh 2× YTA broth and incubated at 37°C with agitation. Protein expression was induced with 0.1 mM IPTG (isopropyl 1-thio-β-D-galactopyranoside) at mid-exponential phase (OD_{600} of 0.6) for CD0757 or at late-exponential phase (OD_{600} of 1.2) for CD1420. The cultures were grown for an additional 4 h at 37°C for CD0757 or 2 h at 25°C for CD0757. Cells were collected by centrifugation, re-suspended in PBS containing 1% Triton X-100 and protease inhibitors (Protease Inhibitor Cocktail, Sigma), and lysed by sonication. CD0757 and CD1420 were recovered by affinity chromatography using the GST purification module (GE Healthcare) with the PreScission protease (GE Healthcare) according to the manufacturer's instructions. After elution, proteins samples were dialyzed against the conservation buffer (50 mM Tris-HCl pH 7.8, 250 mM NaCl, 25 mM KCl, 10 mM $MgCl_2$, 30% glycerol) for 18 h in D-Tube Dialyzer Maxi (MWCO 12–14, Novagen), concentrated by centrifugation on Amicon Ultra-15 columns (MWCO 10, Millipore), and stored at −20°C. Protein concentration was estimated using a BCA Protein Assay Kit (Thermo-Scientific) and purity was determined by SDS-PAGE analysis.

Assays for enzymatic activity and TLC analysis

Diguanylate cyclase and phosphodiesterase activities were measured according to previously described procedures [32,37] with the following modifications. Diguanylate cyclase assays were performed with approximately 1–2 μg of purified proteins in a final volume of 50 μl. Reaction mixtures were pre-incubated for 5 min at 30°C in the reaction buffer (50 mM Tris-HCl pH 7.8, 250 mM NaCl, 25 mM KCl, 10 mM $MgCl_2$). DGC reactions were initiated by adding 33.3 nM [α-^{33}P]-GTP (0.1 μCi μl^{-1}) and incubated at 30°C. Samples were taken at various times, and the reactions were stopped by addition of one volume 0.5 M EDTA. Radiolabeled c-di-GMP for phosphodiesterase activity assays was synthesized using purified DgcK [32]. Purified DgcK (30 μg) was incubated 8 h at 30°C in the reaction buffer to completely convert [α-^{33}P]-GTP into c-di-GMP. Reactions were stopped by denaturing at 99°C for 15 min, centrifuged for 2 min at 16,000 g to eliminate DgcK and recover the supernatant containing the radiolabeled c-di-GMP. Phosphodiesterase assays were performed with approximately 1–2 μg of purified proteins in a final volume of 50 μl of reaction buffer containing 20 nM prepared radiolabeled

c-di-GMP (0.1 μCi μl^{-1}) with or without 100 μM GTP. One unit of snake venom phosphodiesterase (Phosphodiesterase I, Worthington) suspended in SVPD conservation buffer (100 mM Tris-HCl pH 8.0, 100 mM NaCl, 14 mM $MgCl_2$, 50% glycerol) was used as a positive control in PDEA assays. Proteins denatured at 99°C for 15 min were used as negative controls in both DGC and PDEA assays. Reaction products were analyzed by TLC as described before [32]. Briefly aliquots (2–4 μl) were spotted on polyethyleneimine-cellulose TLC plates (Sigma) previously washed in 0.5 M LiCl and air dried. Plates were then soaked for 5 min in methanol, dried, and developed in 2:3 (v/v) saturated $(NH_4)_2SO_4$/1.5 M KH_2PO_4 (pH 3.5). Plates were allowed to dry prior to exposition to a phosphor imaging screen (Molecular Dynamics). Data were collected and analyzed using a FX molecular imager and the Quantity One software (Bio-Rad).

Supporting Information

Figure S1 Domain composition and organization of the 3 putative c-di-GMP-signalling proteins encoded by *C. difficile* strains other than strain 630. Domain composition and organization of the proteins are depicted essentially as in Figure 1. The HD-GYP domain is colored in orange.
(TIF)

Figure S2 Analysis of intracellular c-di-GMP by HPLC. 10 μM synthetic c-di-GMP (A), nucleotide extracts from *V. cholerae* N16961 expressing *lacZ* from pBAD-TOPO/lacZ/V5-His (negative control) (B), the DGC N-448 from pCD0522-(N-448) (C), or the DGC CD1420 from pCD1420 (D). Spectra of synthetic c-di-GMP (E), and of peaks with corresponding retention times in the nucleotide extracts from *V. cholerae* N16961 expressing the DGC N-448 (F) or the DGC CD1420 (G). 100 mg of cells were used for all nucleotide extractions.
(TIF)

Figure S3 Phylogenetic tree of selected Clostridiaceae species. The tree was generated by alignment of the *rpoB* sequences using the neighbor-joining method. The support for each branch is indicated by the value at each node (in percent) as determined by 1,000 bootstrap samples. Only values ≥70% are shown. *Staphylococcus aureus* COL and *Bacillus subtilis* 168 were taken as outgroup strains. Colored boxes highlight the *Clostridium difficile* strains (cyan), the two Clostridiaceae strains with the most c-di-GMP regulatory proteins (yellow) and the two Clostridiaceae strains with the least c-di-GMP regulatory proteins (magenta). Total c-di-GMP regulatory proteins per species are shown in parenthesis next to colored boxes.
(TIF)

Table S1 Predicted and observed activities of the 31 most conserved putative c-di-GMP turn-over proteins.
(DOC)

Table S2 Accession numbers of genomes of *C. difficile* strains used in this study.
(DOC)

Table S3 Clostridiaceae-encoded proteins containing GGDEF, EAL, or HD-GYP domains.
(DOC)

Table S4 Bacterial strains and plasmids used in this study.
(DOC)

Table S5 DNA sequences of the oligonucleotides used in this study.
(DOC)

Text S1 Supplemental Materials and Methods.
(DOC)

Acknowledgments

Author Contributions

Conceived and designed the experiments: EB VB. Performed the experiments: EB VB. Analyzed the data: EB VB. Contributed reagents/materials/analysis tools: VB L-CF FM. Wrote the paper: EB L-CF FM VB.

References

1. Gerding DN (2010) Global epidemiology of *Clostridium difficile* infection in 2010. Infect Control Hosp Epidemiol 31 Suppl 1: S32–34.
2. Kuijper EJ, Coignard B, Tull P (2006) Emergence of *Clostridium difficile*-associated disease in North America and Europe. Clin Microbiol Infect 12 Suppl 6: 2–18.
3. Loo VG, Poirier L, Miller MA, Oughton M, Libman MD, et al. (2005) A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. N Engl J Med 353: 2442–2449.
4. Jenal U, Malone J (2006) Mechanisms of cyclic-di-GMP signaling in bacteria. Annu Rev Genet 40: 385–407.
5. Romling U, Amikam D (2006) Cyclic di-GMP as a second messenger. Curr Opin Microbiol 9: 218–228.
6. Aldridge P, Paul R, Goymer P, Rainey P, Jenal U (2003) Role of the GGDEF regulator PleD in polar development of *Caulobacter crescentus*. Mol Microbiol 47: 1695–1708.
7. Paul R, Weiser S, Amiot NC, Chan C, Schirmer T, et al. (2004) Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. Genes Dev 18: 715–727.
8. Kim YR, Lee SE, Kim CM, Kim SY, Shin EK, et al. (2003) Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. Infect Immun 71: 5461–5471.
9. Kulasakara H, Lee V, Brencic A, Liberati N, Urbach J, et al. (2006) Analysis of *Pseudomonas aeruginosa* diguanylate cyclases and phosphodiesterases reveals a role for bis-(3′-5′)-cyclic-GMP in virulence. Proc Natl Acad Sci U S A 103: 2839–2844.
10. Merkel TJ, Stibitz S, Keith JM, Leef M, Shahin R (1998) Contribution of regulation by the *bvg* locus to respiratory infection of mice by *Bordetella pertussis*. Infect Immun 66: 4367–4373.
11. Tischler AD, Camilli A (2005) Cyclic diguanylate regulates *Vibrio cholerae* virulence gene expression. Infect Immun 73: 5873–5882.
12. Hecht GB, Newton A (1995) Identification of a novel response regulator required for the swarmer-to-stalked-cell transition in *Caulobacter crescentus*. J Bacteriol 177: 6223–6229.
13. Galperin MY, Natale DA, Aravind L, Koonin EV (1999) A specialized version of the HD hydrolase domain implicated in signal transduction. J Mol Microbiol Biotechnol 1: 303–305.
14. Tal R, Wong HC, Calhoon R, Gelfand D, Fear AL, et al. (1998) Three *cdg* operons control cellular turnover of cyclic di-GMP in *Acetobacter xylinum*: genetic organization and occurrence of conserved domains in isoenzymes. J Bacteriol 180: 4416–4425.
15. Romling U, Gomelsky M, Galperin MY (2005) C-di-GMP: the dawning of a novel bacterial signalling system. Mol Microbiol 57: 629–639.
16. Holland LM, O'Donnell ST, Ryjenkov DA, Gomelsky L, Slater SR, et al. (2008) A staphylococcal GGDEF domain protein regulates biofilm formation independently of cyclic dimeric GMP. J Bacteriol 190: 5178–5189.
17. Shang F, Xue T, Sun H, Xing L, Zhang S, et al. (2009) The *Staphylococcus aureus* GGDEF domain-containing protein, GdpS, influences protein A gene expression in a cyclic diguanylic acid-independent manner. Infect Immun 77: 2849–2856.
18. Benach J, Swaminathan SS, Tamayo R, Handelman SK, Folta-Stogniew E, et al. (2007) The structural basis of cyclic diguanylate signal transduction by PilZ domains. EMBO J 26: 5153–5166.
19. Paul K, Nieto V, Carlquist WC, Blair DF, Harshey RM (2010) The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a "backstop brake" mechanism. Mol Cell 38: 128–139.
20. Ryjenkov DA, Simm R, Romling U, Gomelsky M (2006) The PilZ domain is a receptor for the second messenger c-di-GMP: the PilZ domain protein YcgR controls motility in enterobacteria. J Biol Chem 281: 30310–30314.
21. Krasteva PV, Fong JC, Shikuma NJ, Beyhan S, Navarro MV, et al. (2010) *Vibrio cholerae* VpsT regulates matrix production and motility by directly sensing cyclic di-GMP. Science 327: 866–868.
22. Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, et al. (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. Science 321: 411–413.
23. Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR (2010) An allosteric self-splicing ribozyme triggered by a bacterial second messenger. Science 329: 845–848.
24. Galperin MY, Higdon R, Kolker E (2010) Interplay of heritage and habitat in the distribution of bacterial signal transduction systems. Mol Biosyst 6: 721–728.
25. Finn RD, Mistry J, Tate J, Coggill P, Heger A, et al. (2010) The Pfam protein families database. Nucleic Acids Res 38: D211–222.
26. Christen M, Christen B, Folcher M, Schauerte A, Jenal U (2005) Identification and characterization of a cyclic di-GMP-specific phosphodiesterase and its allosteric control by GTP. J Biol Chem 280: 30829–30837.
27. Kumar M, Chatterji D (2008) Cyclic di-GMP: a second messenger required for long-term survival, but not for biofilm formation, in *Mycobacterium smegmatis*. Microbiology 154: 2942–2955.
28. Rao F, Yang Y, Qi Y, Liang ZX (2008) Catalytic mechanism of cyclic di-GMP-specific phosphodiesterase: a study of the EAL domain-containing RocR from *Pseudomonas aeruginosa*. J Bacteriol 190: 3622–3631.
29. Wassmann P, Chan C, Paul R, Beck A, Heerklotz H, et al. (2007) Structure of BeF$_3^-$ modified response regulator PleD: implications for diguanylate cyclase activation, catalysis, and feedback inhibition. Structure 15: 915–927.
30. Ulrich LE, Zhulin IB (2010) The MiST2 database: a comprehensive genomics resource on microbial signal transduction. Nucleic Acids Res 38: D401–407.
31. Beyhan S, Tischler AD, Camilli A, Yildiz FH (2006) Transcriptome and phenotypic responses of *Vibrio cholerae* to increased cyclic di-GMP level. J Bacteriol 188: 3600–3613.
32. Bordeleau E, Brouillette E, Robichaud N, Burrus V (2010) Beyond antibiotic resistance: integrating conjugative elements of the SXT/R391 family that encode novel diguanylate cyclases participate to c-di-GMP signalling in *Vibrio cholerae*. Environ Microbiol 12: 510–523.
33. Kirillina O, Fetherston JD, Bobrov AG, Abney J, Perry RD (2004) HmsP, a putative phosphodiesterase, and HmsT, a putative diguanylate cyclase, control Hms-dependent biofilm formation in *Yersinia pestis*. Mol Microbiol 54: 75–88.
34. Christen B, Christen M, Paul R, Schmid F, Folcher M, et al. (2006) Allosteric control of cyclic di-GMP signaling. J Biol Chem 281: 32015–32024.
35. Malone JG, Williams R, Christen M, Jenal U, Spiers AJ, et al. (2007) The structure-function relationship of WspR, a *Pseudomonas fluorescens* response regulator with a GGDEF output domain. Microbiology 153: 980–994.
36. Hickman JW, Tifrea DF, Harwood CS (2005) A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. Proc Natl Acad Sci U S A 102: 14422–14427.
37. Tamayo R, Tischler AD, Camilli A (2005) The EAL domain protein VieA is a cyclic diguanylate phosphodiesterase. J Biol Chem 280: 33324–33330.
38. Seshasayee AS, Fraser GM, Luscombe NM (2010) Comparative genomics of cyclic-di-GMP signalling in bacteria: post-translational regulation and catalytic activity. Nucleic Acids Res.
39. Mohan R, Namsolleck P, Lawson PA, Osterhoff M, Collins MD, et al. (2006) *Clostridium asparagiforme* sp. nov., isolated from a human faecal sample. Syst Appl Microbiol 29: 292–299.
40. Song Y, Liu C, Molitoris DR, Tomzynski TJ, Lawson PA, et al. (2003) *Clostridium bolteae* sp. nov., isolated from human sources. Syst Appl Microbiol 26: 84–89.
41. Tagliabue L, Antoniani D, Maciag A, Bocci P, Raffaelli N, et al. (2010) The diguanylate cyclase YddV controls production of the exopolysaccharide poly-N-acetylglucosamine (PNAG) through regulation of the PNAG biosynthetic pgaABCD operon. Microbiology 156: 2901–2911.
42. Garcia B, Latasa C, Solano C, Garcia-del Portillo F, Gamazo C, et al. (2004) Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. Mol Microbiol 54: 264–277.
43. Zogaj X, Nimtz M, Rohde M, Bokranz W, Romling U (2001) The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. Mol Microbiol 39: 1452–1463.
44. Tamayo R, Schild S, Pratt JT, Camilli A (2008) Role of cyclic Di-GMP during el tor biotype *Vibrio cholerae* infection: characterization of the in vivo-induced cyclic Di-GMP phosphodiesterase CdpA. Infect Immun 76: 1617–1627.
45. Tasteyre A, Barc MC, Collignon A, Boureau H, Karjalainen T (2001) Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. Infect Immun 69: 7937–7940.
46. Vijay-Kumar M, Gewirtz AT (2009) Flagellin: key target of mucosal innate immunity. Mucosal Immunol 2: 197–205.

47. Schirmer T, Jenal U (2009) Structural and mechanistic determinants of c-di-GMP signalling. Nat Rev Microbiol 7: 724–735.

48. Duerig A, Abel S, Folcher M, Nicollier M, Schwede T, et al. (2009) Second messenger-mediated spatiotemporal control of protein degradation regulates bacterial cell cycle progression. Genes Dev 23: 93–104.

49. Dineen SS, McBride SM, Sonenshein AL (2010) Integration of metabolism and virulence by *Clostridium difficile* CodY. J Bacteriol 192: 5350–5362.

50. Ratnayake-Lecamwasam M, Serror P, Wong KW, Sonenshein AL (2001) *Bacillus subtilis* CodY represses early-stationary-phase genes by sensing GTP levels. Genes Dev 15: 1093–1103.

51. Shivers RP, Sonenshein AL (2004) Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. Mol Microbiol 53: 599–611.

52. Dineen SS, Villapakkam AC, Nordman JT, Sonenshein AL (2007) Repression of *Clostridium difficile* toxin gene expression by CodY. Mol Microbiol 66: 206–219.

53. Heap JT, Cartman ST, Kuehne SA, Cooksley C, Minton NP (2010) ClosTron-targeted mutagenesis. Methods Mol Biol 646: 165–182.

54. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, et al. (2010) The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. J Microbiol Methods 80: 49–55.

55. Heap JT, Pennington OJ, Cartman ST, Minton NP (2009) A modular system for *Clostridium* shuttle plasmids. J Microbiol Methods 78: 79–85.

56. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.

57. Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35: D61–65.

58. Lukashin AV, Borodovsky M (1998) GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res 26: 1107–1115.

59. Kall L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction–the Phobius web server. Nucleic Acids Res 35: W429–432.

60. Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. Science 252: 1162–1164.

61. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.

62. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics 25: 1189–1191.

63. Dower WJ, Miller JF, Ragsdale CW (1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res 16: 6127–6145.

64. Occhino DA, Wyckoff EE, Henderson DP, Wrona TJ, Payne SM (1998) *Vibrio cholerae* iron transport: haem transport genes are linked to one of two sets of *tonB*, *exbB*, *exbD* genes. Mol Microbiol 29: 1493–1507.