Cell Cycle Constraints and Environmental Control of Local DNA Hypomethylation in α-Proteobacteria

Short Title: Control of DNA Hypomethylation in α-Proteobacteria

Silvia Ardissone[1]*, Peter Redder[1#a], Giancarlo Russo[2], Antonio Frandi[1,#b], Coralie Fumeaux[1, #c], Andrea Patrignani[2], Ralph Schlapbach[2], Laurent Falquet[3] and Patrick H. Viollier[1]*

[1] Department of Microbiology and Molecular Medicine, Institute of Genetics & Genomics in Geneva (iGE3), Faculty of Medicine, University of Geneva, Geneva, Switzerland.

[2] Functional Genomics Center Zurich, ETH/University of Zürich, Zürich, Switzerland.

[3] Biochemistry Unit, Dept. of Biology, University of Fribourg and Swiss Institute of Bioinformatics, Fribourg, Switzerland.

[#a] Current address: LMGM, Centre de Biologie Intégrative, Université Paul Sabatier, Toulouse, France.

[#b] Current address: Department of Fundamental Microbiology, University of Lausanne, Lausanne, Switzerland.

[#c] Current address: Department of Microbiology and Immunology, Harvard Medical School, Boston, MA, 02115, USA.

* Corresponding authors

Silvia.Ardissone@unige.ch (SA)

Patrick.Viollier@unige.ch (PV)

Abstract

Heritable DNA methylation imprints are ubiquitous and underlie genetic variability from bacteria to humans. In microbial genomes, DNA methylation has been implicated in gene transcription, DNA replication and repair, nucleoid segregation, transposition and virulence of pathogenic strains. Despite the importance of local (hypo)methylation at specific loci, how and when these patterns are established during the cell cycle remains poorly characterized. Taking advantage of the small genomes and the synchronizability of α-proteobacteria, we discovered that conserved determinants of the cell cycle transcriptional circuitry establish specific hypomethylation patterns in the cell cycle model system *Caulobacter crescentus*. We used genome-wide methyl-N6-adenine (m6A-) analyses by restriction-enzyme-cleavage sequencing (REC-Seq) and single-molecule real-time (SMRT) sequencing to show that MucR, a transcriptional regulator that represses virulence and cell cycle genes in S-phase but no longer in G1-phase, occludes 5'-GANTC-3' sequence motifs that are methylated by the DNA adenine methyltransferase CcrM. Constitutive expression of CcrM or heterologous methylases in at least two different α-proteobacteria homogenizes m6A patterns even when MucR is present and affects promoter activity. Environmental stress (phosphate limitation) can override and reconfigure local hypomethylation patterns imposed by the cell cycle circuitry that dictate when and where local hypomethylation is instated.

Author Summary

DNA methylation is the post-replicative addition of a methyl group to a base by a methyltransferase that recognise a specific sequence, and represents an epigenetic regulatory mechanism in both eukaryotes and prokaryotes. In microbial genomes, DNA methylation has been implicated in gene transcription, DNA replication and repair, nucleoid segregation, transposition and virulence of pathogenic strains. CcrM is a conserved, cell cycle regulated adenine methyltransferase that methylates GANTC sites in α-proteobacteria. N^6-methyl-adenine (m6A) patterns generated by

CcrM can change the affinity of a given DNA-binding protein for its target sequence, and therefore affect gene expression. Here, we combine restriction enzyme cleavage-deep sequencing (REC-Seq) with SMRT sequencing to identify hypomethylated 5'-GANTC-3' (GANTCs) in α-proteobacterial genomes instated by conserved cell cycle factors. By comparing SMRT and REC-Seq data with chromatin immunoprecipitation-deep sequencing data (ChIP-Seq) we show that a conserved transcriptional regulator, MucR, induces local hypomethylation patterns by occluding GANTCs to the CcrM methylase and we provide evidence that this competition occurs during S-phase, but not in G1-phase cells. Furthermore, we find that environmental signals (such as phosphate depletion) are superimposed to the cell cycle control mechanism and can override the specific hypomethylation pattern imposed by the cell cycle transcriptional circuitry.

Keywords

REC-Seq; SMRT sequencing; hypomethylation; adenine methylation; CcrM; MucR; α-proteobacteria; *Caulobacter crescentus*; *Sinorhizobium meliloti*; phosphate limitation

Introduction

DNA methylation is a conserved epigenetic modification that occurs from bacteria to humans and is implicated in control of transcription, DNA replication/repair, innate immunity and pathogenesis [1, 2]. Originally described as a mechanism that protects bacteria from invading foreign (viral) DNA [3], methyl-N6-adenine (m6A) modifications are thought to direct infrequent and stochastic phenotypic heterogeneity in bacterial cells [4, 5] and were recently implicated in transcriptional control of lower eukaryotic genomes [4, 5] and silencing in mouse embryonic stem cells [6-8].

How local changes in methylation are instated during the cell cycle remains poorly explored, even in γ-proteobacteria such as *Escherichia coli* and *Salmonella enterica*, as cell cycle studies on cell populations are cumbersome and require genetic manipulation [9]. Moreover, the replication regulator SeqA that controls the methylation state by preferentially binding hemi-methylated sequences is only encoded in γ-proteobacteria, suggesting that other mechanisms are likely operational in other systems [9, 10]. Model systems in which cell populations can be synchronized without genetic intervention are best suited to illuminate the interplay between methylation and cell cycle [11, 12]. The fresh-water bacterium *Caulobacter crescentus* and more recently the plant symbiont *Sinorhizobium meliloti* that reside in distinct environmental niches are such cell cycle model systems [13]. Akin to other α-proteobacteria, *C. crescentus* and *S. meliloti* divide asymmetrically into a smaller G1-phase cell and a larger S-phase cell and use conserved transcriptional regulators arranged in modules to coordinate transcription with cell cycle progression [13-16] (Fig 1A). In *C. crescentus*, MucR1 and MucR2 were recently shown to negatively regulate numerous promoters that are activated by the cell cycle transcriptional regulator A (CtrA) in G1-phase. MucR orthologs control virulence functions in α-proteobacterial pathogens and symbionts, but can also control cell cycle-regulated promoters in *C. crescentus* [17-20]. MucR1/2 target promoters by way of an ancestral zinc finger-like fold and both proteins are present throughout the *C. crescentus* cell cycle [17, 21, 22] (Fig 1A). By contrast, the OmpR-like DNA-binding response regulator CtrA is activated by phosphorylation and is only present in G1 and late S-phase cells [23, 24], but not in early S-phase cells (Fig 1A). The promoter controlling

93 expression of the conserved DNA methyltransferase CcrM is among the targets activated by
94 phosphorylated CtrA (CtrA~P) in late S-phase [15, 17, 25-27]. CcrM introduces m6A marks at sites
95 harbouring the recognition sequence 5'-GANTC-3' (henceforth GANTCs) once passage of the DNA
96 replication fork leaves GANTCs hemi-methylated (Fig 1B). CcrM is an unstable protein degraded by
97 the ATP-dependent protease Lon throughout the cell cycle [28, 29]. Since the *ccrM* gene is expressed
98 only in late S-phase cells, the time of expression dictates when the unstable CcrM protein is present
99 during the cell cycle. CcrM no longer cycles when it is expressed from a constitutive promoter in
100 otherwise *WT* cells or when Lon is inactivated [28, 30].

101 **Fig 1. Regulation of *C. crescentus* cell cycle and methylation of the chromosome.**

102 (A) Schematic of the *C. crescentus* cell cycle and the regulatory interactions that control G1-phase
103 promoters. Transcription from G1-phase promoters is activated by phosphorylated CtrA (CtrA~P, in
104 blue) and repressed by MucR1/2. CtrA~P accumulates in pre-divisional and swarmer (SW) cells, and
105 is eliminated by regulated proteolysis in the stalked (ST) cell upon compartmentalization and at the
106 swarmer to stalked cell differentiation. The red bar indicates that MucR1/2 proteins are present at all
107 stages of the cell cycle. (B) Schematic of the chromosome replication and adenine methylation (m6A)
108 during the *C. crescentus* cell cycle. Chromosome replication and methylation are shown for the same
109 cell cycle stages depicted in panel A. In non-replicative SW cells (T10) the chromosome is methylated
110 on both strands. Upon differentiation into ST cells (T40), chromosome replication start from the origin
111 (blue star) located at the old cell pole; progression of the replisomes (orange dots) generates hemi-
112 methylated chromosomes (black and green lines represent respectively the methylated and
113 unmethylated strand; T70 and T100). The cell cycle-regulated adenine methyltransferase CcrM is
114 synthesized only in late pre-divisional cells, where it methylates the newly synthesized chromosome
115 strands. CcrM is then specifically proteolyzed by the Lon protease. The purple numbers indicate the
116 position along the chromosome of the *P169* (1), *P1149* (2) and *P2901* (3) promoters. T10, T40, T70
117 and T100 indicate the time after synchronization at which the samples for anti-MucR1 ChIP-Exo were
118 taken (see text).

With the advent of SMRT (single-molecule real-time) sequencing it is now possible to obtain m6A-methylome information of bacterial genomes at single base pair resolution [31, 32]. A recent cell cycle methylome analysis of *C. crescentus* by SMRT-sequencing revealed the large majority of GANTCs switch from hemi-methylated to a full methylated state (m6A-marked GANTCs on both strands) at the onset of CcrM expression [12]. Interestingly, a few sites were consistently hypomethylated, indicating that site-specific mechanisms control local hypomethylation patterns. Local hypomethylation patterns may arise if specific DNA-binding proteins and/or restricted local chromosome topology block access of CcrM to such GANTCs. Here, we combine restriction enzyme cleavage-deep sequencing (REC-Seq) with SMRT sequencing to unearth hypomethylated GANTCs in the genomes of wild type (*WT*) and mutant *C. crescentus* and *S. meliloti*. We show that the conserved transcriptional regulator MucR induces local m6A-hypomethylation by preventing CcrM from accessing GANTCs during S-phase, but only when CcrM cycles. Since repression of MucR target promoters is normally overcome in G1-phase, our data suggest that MucR is unable to shield GANTCs when CcrM is artificially present in G1 cells. Lastly, we discovered that phosphate starvation promotes methylation of specific MucR-shielded GANTCs, revealing an environmental override of the control system that normally instates local hypomethylation patterns during the cell cycle.

Results

Identification and analysis of hypomethylated GANTCs by restriction enzyme cleavage (REC-Seq).

Detection of hypomethylated sites by SMRT-sequencing requires sufficient sequencing depth and sophisticated bioinformatic analysis to differentiate unmethylated GANTCs from methylated ones. Since unmethylated GANTCs can be conveniently enriched for in *C. crescentus* by restriction enzyme cleavage using the *Hin*fI restriction enzyme (which only cleaves unmethylated GANTCs)

144 [33], we sought to apply *Hin*fI-based cleavage followed by Illumina-based deep-sequencing (REC-
145 Seq) to identify hypomethylated GANTCs, similar to a previous procedure used for analysis of
146 hypomethylated m6A sites in the unrelated γ-proteobacterium *Vibrio cholerae* [34]. We tested REC-
147 Seq on *Hin*fI-treated genomic DNA (gDNA) from *C. crescentus* and, following bioinformatic
148 filtering, obtained a list of unprotected GANTCs scaling with *Hin*fI cleavage efficiency ("score" in S1
149 Table). Since nearly all GANTCs suggested to be consistently unmethylated by SMRT sequencing
150 [12] are represented as high scoring GANTCs in the REC-Seq (note the medium differences or limited
151 SMRT sequencing depth may explain the differences), we concluded that REC-Seq captures
152 hypomethylated GANTCs in scaling manner (see below where selected sites cleaved in *WT* are no
153 longer cleaved in the Δ*mucR1/2* mutant). Since CcrM also methylates GANTCs in other α-
154 proteobacteria [35, 36], we also determined the hypomethylated GANTCs on the multipartite genome
155 of *S. meliloti* [37] by *Hin*fI REC-Seq and found such hypomethylated sites on the chromosome and
156 both megaplasmids (S1 Table).

157 To validate the *Hin*fI REC-Seq approach, we conducted REC-Seq (using the methylation-
158 sensitive *Mbo*I restriction enzyme) on gDNA from *Escherichia coli* K12 and *V. cholerae*, as
159 previously determined either by SMRT sequencing or REC-Seq [10, 34]. The Dam methylase
160 introduces m6A marks at GATCs in many γ-proteobacterial genomes [4] that protect from cleavage by
161 *Mbo*I. As known unmethylated sites in these control experiments indeed emerged with high score (S2
162 Table), we conclude that *Hin*fI REC-Seq is an efficient method to detect and quantitate GANTCs that
163 escape methylation by CcrM.

164

165 **Several hypomethylated GANTCs in *C. crescentus* are MucR target sites.**

166 Having identified hypomethylated GANTCs in the *C. crescentus* genome by *Hin*fI REC-Seq,
167 we noted that many high scoring GANTCs lie in regions that are occupied by MucR1/2 as determined
168 by previous chromatin-immunoprecipitation deep-sequencing (ChIP-Seq) analysis [17]. Of the hits
169 with a score higher than 100, one third lie in MucR1/2 target sequences, and the proportion is even

higher (50%) in the case of the 50 top hits (Table 1 and S1 Table). To test if MucR1/2 occludes these GANTCs from methylation by CcrM, we conducted *Hin*fI-cleavage analysis of gDNA from *WT* (NA1000) and Δ*mucR1/2* double mutant by qPCR (henceforth *Hin*fI-qPCR assay) at six MucR1/2 target sites. The *CCNA_00169* promoter (henceforth P*169*) contains four GANTCs; the *CCNA_02901* promoter (P*2901*), the *CCNA_01149* promoter (P*1149*) and the *CCNA_01083* internal sequence contain two GANTCs each; the *CCNA_02830* and *CCNA_03248* promoters (P*2830* and P*3248*) carry one GANTC each (Fig 2A). A high percentage (100%) of methylation in the *Hin*fI-qPCR assay indicates that *Hin*fI cannot cleave this site because of prior methylation by CcrM, whereas a low percentage reflects efficient cleavage of the non-methylated DNA by *Hin*fI. In *WT* gDNA these six MucR1/2-target sequences are almost completely cleaved by *Hin*fI, indicating that the GANTCs are hypomethylated in the presence of MucR1/2. However, these sites are methylated and therefore not cleaved by *Hin*fI in Δ*mucR1/2* cells (Fig 2B). As control for the specificity of the *Hin*fI-qPCR assay we conducted the same analysis on sequences that are not MucR1/2 targets harbouring either i) a hypomethylated GANTC (P_{nagA}), ii) several methylated GANTCs (P_{podJ}) or iii) a control sequence that does not contain GANTCs (P_{xylX}). These controls revealed a level of amplification in the *Hin*fI-qPCR assay as predicted (S1A Fig) and showed no difference between *WT* and Δ*mucR1/2* cells. Thus, only hypomethylated sequences that are bound by MucR1/2 *in vivo* are converted to methylated GANTCs in the absence of MucR1/2.

Fig 2. MucR occludes specific GANTC sites from methylation.

(A) Schematic of the loci carrying hypomethylated GANTCs occluded by MucR. The position of the hypomethylated GANTCs identified by Kozdon et al. [12] is indicated by purple asterisks. Red lines represent the occupancy of MucR1 and the values ($x10^4$ *per-base* coverage) calculated by the super-resolution bioinformatic approach represent the average of the four time points (T10, T40, T70 and T100, as described in the Methods). The MucR-dependent transcription start sites, determined by TSS-EMOTE, are indicated by black (sense) and green (antisense) arrows. Dashed arrows indicate transcription start sites found in both *WT* and Δ*mucR1/2* strains (*CCNA_01083-CCNA_01084*) or down-regulated in the Δ*mucR1/2* strain compared to the *WT* (promoter of *CCNA_02831*). (B)

Methylation percentage of the loci shown in panel A in the *WT* and Δ*mucR1/2* strains, as determined by *Hin*fI-qPCR.

SMRT sequencing of *WT* and Δ*mucR1/2* gDNA supported the result that these GANTCs carry m6A marks as inferred by a high characteristic interpulse-duration (IPD) ratio observed in Δ*mucR1/2* versus *WT* cells (S3 Table). Interestingly, this analysis also revealed eleven GANTCs with the inverse behaviour, i.e. a low IPD ratio in Δ*mucR1/2* versus *WT* cells, suggesting that they no longer carry m6A marks in the absence of MucR1/2. To confirm this result we conducted *Hin*fI-qPCR assays at two of these GANTCs: the *CCNA_01248* promoter (P*1248*) and the *CCNA_03426* promoter (P*3426*). As predicted by the methylome analysis, we observed that the methylation percentage of these GANTCs was reduced in Δ*mucR1/2* versus *WT* (S1B Fig). On the basis of these experiments, we conclude that MucR1/2 prevents m6A-methylation by CcrM at several MucR1/2-target sequences, but can also facilitate methylation at other sites. This would likely occur by an indirect mechanism involving other MucR-dependent DNA-binding proteins that compete with CcrM at certain GANTCs.

To obtain a global picture of hypomethylated GANTCs in the absence of MucR1/2, we conducted REC-Seq analysis on gDNA extracted from the Δ*mucR1/2* strain (Table 1 and S1 Table). Comparison of the REC-Seq data for *WT* and Δ*mucR1/2* cells (S2 Fig) supported the conclusion that binding of MucR1/2 prevents methylation by CcrM, as the GANTCs tested by *Hin*fI-qPCR (shown in Fig 2) have a high REC-Seq score in *WT* and a low REC-Seq score (or they are not detected) in the Δ*mucR1/2* strain. Moreover, most of the GANTCs that show a strong decrease in score between *WT* and Δ*mucR1/2* cells are also lying in regions directly bound by MucR1/2 (Table 1 and S1 Table), based on ChIP-Exo (S4 Table) and published ChIP-Seq data [17].

Table 1. **REC-Seq in *WT* and Δ*mucR1/2*.**

GANTC sites with the highest (top 50) REC-Seq score in *WT C. crescentus* are listed. The complete list of REC-Seq data for both *WT* and Δ*mucR1/2* strains is reported in Supplementary Table S1.

Position of GANTC site		REC-Seq Score in		Possible association	[1]Persistently unmethylated	[2]MucR target
		WT	*ΔR1R2*			
180973	180975	17756	1802	CCNA_00169		R1, R2
3781337	3781339	15184	8230	CCNA_R0079		
3992382	3992384	10934	5474	CCNA_03826/CCNA_03827		
738031	738033	10236	4678	CCNA_00681/CCNA_00682	Y	
3415859	3415861	9733	91	CCNA_03248	Y	R1, R2
2321619	2321621	9628	5563	CCNA_02167	Y	
2509257	2509259	9166	4171	CCNA_02366/CCNA_02367	Y	
1014421	1014423	8697	4155	CCNA_R0125/CCNA_00938/ CCNA_00939	Y	
1253987	1253989	7236	0	CCNA_01149	Y	R1, R2
1188431	1188433	7162	0	CCNA_01083	Y	R1
3431586	3431588	6912	116	CCNA_03262/CCNA_03263		R1, R2
3882781	3882783	6740	1061	CCNA_R0090/CCNA_03717/ CCNA_03718		
479267	479269	6634	14	CCNA_00466		R1, R2
458563	458565	6417	3	CCNA_00451	Y	
461814	461816	6211	4061	CCNA_00454/CCNA_03920	Y	
2086467	2086469	5523	19	CCNA_01944/CCNA_R0152		R1, R2
435741	435743	5173	0	CCNA_00425	Y	R1, R2
2056538	2056540	4840	2243	CCNA_01912/CCNA_01913		
1544245	1544247	4824	0	CCNA_01426	Y	R1, R2
1645085	1645087	4804	1875	CCNA_R0040		
1375838	1375840	4795	2782	CCNA_01247/CCNA_01248		
626625	626627	4669	3281	CCNA_00594/CCNA_R0112		
181127	181129	4639	0	CCNA_0169	Y	R1, R2
2335870	2335872	4636	571	CCNA_03962		
3885153	3885155	4552	247	CCNA_03719/CCNA_03721		
2984830	2984832	3903	0	CCNA_03980/CCNA_02830	Y	R1, R2
1567545	1567547	3901	3	CCNA_01457		R1, R2
1187860	1187862	3825	16	CCNA_01083	Y	R1, R2
2396048	2396050	3680	1531	CCNA_02246/CCNA_02247		
3056349	3056351	3578	0	CCNA_02901	Y	R1, R2
1187837	1187839	3298	7	CCNA_01083	Y	R1, R2
738020	738022	3255	1557	CCNA_00681/CCNA_00682	Y	
3056446	3056448	3249	0	CCNA_02901	Y	R1, R2
1906979	1906981	3099	17	CCNA_01779/CCNA_01780		R1, R2
181231	181233	2934	3	CCNA_00169	Y	R1, R2
3359883	3359885	2887	1682	CCNA_03199		
1254029	1254031	2873	0	CCNA_01149	Y	R1, R2
713385	713387	2823	0	CCNA_00657	Y	R1, R2
181197	181199	2712	0	CCNA_00169	Y	R1, R2

3261049	3261051	2571	283	CCNA_03108		
3188019	3188021	2520	1834	CCNA_03032/CCNA_03033	Y	
3987164	3987166	2451	489	CCNA_R0199		
2889722	2889724	2404	0	CCNA_02726	Y	R1, R2
2591862	2591864	2365	124	CCNA_02452		
2044585	2044587	2309	1765	CCNA_01902		
3592546	3592548	2178	7814	CCNA_03426		
1690112	1690114	2126	1239	CCNA_01573/CCNA_01574		
902058	902060	2098	6	CCNA_00836		R1, R2
2426668	2426670	1824	108	CCNA_02275/CCNA_02277		
1544275	1544277	1765	0	CCNA_01426		R1, R2

(1) Sites labelled with Y were identified as persistently unmethylated during the cell cycle by Kozdon et al. [12]

(2) Indicates whether a given GANTC site overlaps with MucR1 (R1) or MucR2 (R2) peaks identified by ChIP-Exo and ChIP-Seq (S4 Table and [17])

Conditions that impair local GANTC hypomethylation by MucR1/2.

Since CcrM is restricted to late S-phase and MucR1/2-repression is overcome in G1-phase [17, 28], we tested if MucR1/2-bound GANTCs are still hypomethylated when CcrM no longer cycles. To this end we used two strains: the Δ*lon*::Ω (henceforth *lon*) mutant, as the Lon protease is responsible for degradation of CcrM throughout the cell cycle and upon inactivation of Lon the CcrM protein accumulates also in G1-cells, although it is only synthesized in S-phase [28, 29], and a strain with a second copy of the *ccrM* gene under control of the constitutive P_{lac} promoter (integrated at the *ccrM* locus, *ccrM*::P_{lac}-*ccrM*) [30, 33]. Indeed, *Hin*fI-qPCR analysis revealed that the fraction of methylated P*169*, P*1149* and P*2901* GANTCs increases in *lon* mutant and *ccrM*::P_{lac}-*ccrM* strain relative to *WT* cells (Fig 3A).

To exclude that constitutive presence of CcrM simply prevents MucR binding to DNA because CcrM outnumbers and therefore outcompetes MucR, we conducted several control experiments to demonstrate the specificity of the methylation control at these GANTCs. First, immunoblotting experiments revealed that MucR1/2 levels were maintained in the *lon* and *ccrM*::P_{lac}-*ccrM* strains compared to *WT* (Fig 3C). Second, overexpression of either *WT* MucR1 or of an N-terminally extended (dominant-negative) MucR1 variant from P_{van} on a high copy plasmid (pMT335)

[17, 38] did not prevent methylation of P*169*, P*1149* and P*2901* GANTCs in *lon* mutant cells (Fig 3A) or alter CcrM steady-state levels (S1C Fig). Conversely, constitutive expression of CcrM from the same vector (pMT335) in *WT* cells recapitulated the effect on methylation of the P*169*, P*1149* and P*2901* GANTCs (Fig 3B). Similarly, methylases of *Thermoplasma acidophilum* (*TA*), *Helicobacter pylori* (*HP*) or *Haemophilus influenzae* (*Hinf*), which also specifically methylate GANTCs but are not related to α-proteobacterial CcrM, also lead to methylation of these hypomethylated GANTCs when expressed from pMT335 (Fig 3B). By contrast, the methylation state of GANTCs at the *parS* locus was not significantly altered by the expression of the methylases or by the *lon* mutation (S1D Fig). On the basis that CcrM and unrelated methylases are able to compete against MucR1/2 for methylation of P*169*, P*1149* and P*2901* GANTCs when expressed constitutively, we hypothesize that MucR1/2 no longer efficiently compete with CcrM in G1-phase when both proteins are present at this time (Fig 3A, 3B).

Fig 3. Hypomethylation by MucR is impaired in G1-phase cells.

(A) Methylation percentage of the *P169*, *P1149* and *P2901* sequences in strains in which the methyltransferase CcrM is stabilised. The *Hin*fI-qPCR analysis indicates that methylation is increased in cells carrying P_{lac}-*ccrM* or the *lon* mutation. In the case of the *lon* mutant, the methylation of *P169*, *P1149* and *P2901* is not affected by increased levels of MucR1 [*R1*: P_{van}-*mucR1*, *R1 long*: N-terminally extended dominant-negative MucR1 variant expressed from P_{van} on pMT335]. (B) Methylation percentage of the *P169*, *P1149* and *P2901* sequences in *WT* cells that constitutively express *ccrM* or heterologous GANTC-methylases from P_{van} on pMT335 (TA, *Thermoplasma acidophilum*; HP, *Helicobacter pylori*; *hinf*, *Haemophilus influenzae*). (C) Immunoblot showing the steady-state levels of CcrM, MucR1 and MucR2 in *WT* cells as well as the P_{lac}-*ccrM*, *lon* and Δ*mucR1/2* strains. Molecular size standards are indicated on the right as blue lines with the corresponding values in kDa. (D) Genome-wide occupancy of MucR1 in synchronised *WT* cells at four different time points as determined by ChIP-Exo and super-resolution bioinformatic analysis: swarmer (T10), stalked (T40), early pre-divisional (T70) and late pre-divisional cells (T100). The *x* axis represents the nucleotide position on the genome, whereas the *y* axis shows the enrichment ratio

(E.R.) for each promoter region as reported in the S4 Table (see Methods for a detailed description). (E) Heat map of the enrichment ratios of selected loci (those that contain hypomethylated GANTCs occluded by MucR, see Fig 2) at the four time points after synchronization. The *pilA* locus is shown as comparison as it is a well-characterized target of MucR [17]. The heat map shows that MucR1 is constitutively associated with these loci.

To test if MucR1 binds to its targets in G1-phase, we conducted chromatin-immunoprecipitation-followed by deep-sequencing of exonuclease treated fragments (ChIP-Exo), a technique with enhanced resolution compared to conventional ChIP-Seq [39]. We treated with the anti-MucR1 antibody chromatin prepared from synchronized cells at four different time points after synchronization [10 min (T10, G1 phase), 40 min (T40, G1-to-S transition), 70 min (T70, early S-phase), 100 min (T100, late S-phase) (Fig 1B)] and used a bioinformatic algorithm to define the binding sites at super-resolution (see Methods and [40]). Surprisingly, the binding profiles at the four time points appeared to be nearly congruent (Fig 3D) and quantification of the enrichment ratio failed to reveal major changes of MucR1 binding to its targets during the cell cycle (Fig 3E and S4 Table). On the other hand, conformational changes or altered dynamics of binding (i.e. dissociation constants, on- and off-rates) that are undetectable by our methods might allow transcription from the MucR-bound promoters in G1-phase. Transient release of DNA by MucR1/2 or changes in chromatin conformation could provide access to competing DNA binding proteins such as CcrM, RNA polymerase (RNAP) and other transcription factors (like CtrA) in G1-phase to induce methylation or firing of the MucR1/2 target promoters.

MucR-dependent hypomethylation regulates sense and anti-sense transcription.

As MucR1/2 regulates the methylation state of the aforementioned GANTCs, we wondered if the MucR1/2-targets P*169*, P*1149* and P*2901* display promoter activity in a MucR1/2-dependent and/or methylation-dependent manner. To this end, we conducted LacZ (β-galactosidase)-based

promoter probe assays of P*169*-, P*1149*- and P*2901-lacZ* transcriptional reporters (driving expression of a promoterless *lacZ* gene) in *WT* and Δ*mucR1/2* cells and observed that LacZ activity of all reporters was elevated in Δ*mucR1/2* cells versus *WT* (Fig 4A-C). The increase was less dramatic for P*169-lacZ* (156 ± 5.8 % relative to *WT*) than for P*1149*- and P*2901-lacZ* (439 ± 7.4 and 385 ± 40 %, respectively). We then asked if promoter activity is augmented when cycling of CcrM is prevented. Indeed, the P*169*-, P*1149*- and P*2901-lacZ* reporters indicated an increase in promoter activity in the *lon* mutant and P_{lac}-*ccrM* strains compared to *WT* (Fig 4D). Importantly, no increase in LacZ activity was observed in *lon* and P_{lac}-*ccrM* strains with other promoters (P*hvyA* and P*pilA*, Fig 4D) that are bound by MucR1/2 and whose activity is increased in Δ*mucR1/2* cells [17, 41] but contain no hypomethylated GANTCs. We further corroborated these results by showing that constitutive expression of *C. crescentus* CcrM or the *T. acidophilum* GANTC-methylase from P_{van} on pMT335 led to an increase in P*169*-, P*1149*- and P*2901-lacZ* promoter activity (Fig 4E). Consistent with the fact that in Δ*mucR1/2* cells these promoters are no longer hypomethylated, constitutive expression of CcrM from P_{lac}-*ccrM* in Δ*mucR1/2* cells had no significant effect on P*169*-, P*1149*- and P*2901-lacZ* promoter activity (Fig 4E).

Fig 4. MucR and methylation by CcrM regulate transcription from target promoters.

Beta-galactosidase activity of P*169* (*WT* promoter) and P*169** (with all GANTCs mutated to GTNTCs) (A), P*2901* (*WT* promoter) and P*2901** (with the two GANTCs mutated to GTNTCs) (B) and P*1149* (C) in *WT* and Δ*mucR1/2* cells. Mutation of MucR1/2 increases expression from P*169*-, P*2901*- and P*1149-lacZ* independently from the presence of GANTCs. Values are expressed as percentages (activity of *WT* promoter in *WT* cells set at 100%). (D) Beta-galactosidase activity of P*169*-, P*1149*-, P*2901-lacZ* promoter probe constructs and two MucR-dependent control promoter reporters (P_{hvyA}-*lacZ* and P_{pilA}-*lacZ*) in *WT* and cells that constitutively express *ccrM* (ccrM::P_{lac}-*ccrM* or Δ*lon*::Ω). Methylation of the target promoters by CcrM increases the LacZ activity. Values are expressed as percentages (activity in *WT* cells set at 100%). (E) Beta-galactosidase activity of P*169*-, P*1149*- and P*2901-lacZ* in *WT* cells that constitutively express *ccrM* or a heterologous GANTC-methylase from *T. acidophilum* (TA) on plasmid under control of P_{van}. Values are expressed as

322 percentages (activity in *WT* carrying the empty vector set at 100%). (F) Beta-galactosidase activity of
323 P*169*-, P*1149*- and P*2901-lacZ* in *WT* and Δ*mucR1/2* cells that constitutively express *ccrM*
324 (ccrM::P_{lac}-*ccrM*). Values are expressed as percentages (activity in *WT* cells set at 100%).

325

326 To determine if changing the GANTC methylation state (by mutation to GTNTC) in P*169*-
327 and P*2901-lacZ* (5 sites mutated for P*169,* P*169**; 2 sites for P*2901,* P*2901**) also affects promoter
328 activity, we measured LacZ activity of the mutant promoters in *WT* and Δ*mucR1/2* cells and found that
329 they still exhibited MucR1/2-dependency, as the P*169**- and P*2901***-lacZ* were still strongly de-
330 repressed in the absence of MucR1/2 (Fig 4A, 4B). We also observed an increase (136% ± 6%) in
331 activity of P*169***-lacZ* relative to P*169-lacZ* in *WT*, while the activity of P*2901***-lacZ* was decreased
332 compared to P*2901-lacZ.* The mutations may alter the target sequence for other regulator(s) in
333 addition to the methylation properties, thereby affecting transcription directly or indirectly in a
334 positive or negative fashion [42, 43]. For example, P*2901* is bound by the master cell cycle regulator
335 CtrA *in vivo* and the Δ*mucR1/2* mutation is known to affect CtrA expression [17], whereas the P*169*
336 promoter is affected by the phosphate starvation response (see below) [44, 45].

337 LacZ-based assays are a general and indirect measurement of promoter activity, but they do
338 not pinpoint the transcription start sites (TSSs), thus cannot reveal the physical proximity of the TSS
339 relative to the hypomethylated GANTCs. To correlate transcriptional regulation of MucR1/2 and
340 hypomethylated GANTCs, we took advantage of the recently developed RNA-Seq-based strategy for
341 exact mapping of transcriptome ends (EMOTE) [46] that can also be used to map the (unprocessed)
342 5'ends of nascent transcripts that harbour triphosphate 5'end (5'-ppp). To this end, total RNA is first
343 treated with XRN1 (to remove monophosphorylated 5'ends) and then 5'-ppp transcripts are treated
344 with *E. coli* RppH, which converts the 5'ends to a monophosphorylated form that can be ligated to a
345 bar-tagged RNA oligo using T4 RNA ligase [46] (S3 Fig). Comparative TSS-EMOTE analysis in total
346 RNA extracted from *WT* and Δ*mucR1/2* cells unearthed several TSSs that are activated when
347 MucR1/2 is absent (arrows in Fig 2A, S5 Table). Importantly, several of these TSSs were detected in

close proximity to the GANTCs within MucR1/2 target sequences, including P*2901*, P*2830*, P*3248* and *CCNA_01083*. These results, therefore, validate the physical proximity and functional interplay between MucR1/2 and hypomethylated GANTCs. While for weak MucR1/2 target promoters the sequencing depth may have limited their detection by TSS-EMOTE, this analysis unexpectedly revealed several MucR-dependent antisense transcripts with potential regulatory roles (green arrows in Fig 2A). We validated the MucR1/2-dependency of two such antisense promoters (P*2902*_AS and P*3247*_AS) by LacZ-promoter probe assays and detected a substantial increase in activity of P*2902*_AS-*lacZ* and P*3247*_AS-*lacZ* in Δ*mucR1/2* versus *WT* cells (S1E Fig), indicating that these promoters (and the GANTCs within) are clearly MucR1/2 regulated in *C. crescentus*.

Control of hypomethylation of MucR-target promoters in α-proteobacteria

Knowing that MucR is functionally interchangeable in α-proteobacteria [17, 18] and that hypomethylated GANTCs are also detected in the *S. meliloti* multipartite genome by *Hin*fI REC-Seq (see above and S1 Table), we tested whether *S. meliloti* MucR also occludes GANTCs from methylation by CcrM in target promoters. We compared the methylation of *WT* and *mucR*::Tn *S. meliloti* gDNA by *Hin*fI REC-Seq and SMRT-sequencing (S1 and S3 Table). Guided by these data sets, we validated hypomethylation of GANTCs at or near the *SMa1635* (SM2011_RS04470) and *SMa2245* (SM2011_RS06125) genes by *Hin*fI-restriction/qPCR analysis. We chose these GANTCs, located on the symbiotic megaplasmid pSymA, to take advantage of the *S. meliloti* multipartite genome and to explore if MucR-control of hypomethylation also applies to episomal elements such as a symbiotic megaplasmid. *Hin*fI-restriction/qPCR analysis revealed that these GANTCs are largely hypomethylated in *WT* compared to *mucR*::*Tn* cells (Fig 5A, 5B). To confirm that these GANTCs are indeed direct targets of *S. meliloti* MucR, we conducted quantitative ChIP (qChIP) experiments (Fig 5D) with chromatin from *S. meliloti WT* and *mucR*::*Tn* cells precipitated using antibodies to *C. crescentus* MucR2 that recognize *S. meliloti* MucR on immunoblots (S4A Fig). The qChIP experiments revealed that *S. meliloti* MucR indeed binds at or near the hypomethylated *SMa1635* and

374 *SMa2245* GANTCs of *WT* cells (Fig 5D), but not at a control site (*SMc01552*). Moreover, since CcrM
375 is restricted to late S-phase also in *S. meliloti* [14], we tested whether constitutive expression of $ccrM^{Cc}$
376 in *S. meliloti WT* cells affected the methylation of GANTCs at *SMa1635* and *SMa2245*. Ectopic
377 expression of $ccrM^{Cc}$ from P_{lac} on pSRK vector [47] significantly increased the methylation of
378 *SMa1635* and *SMa2245*, showing that *S. meliloti* MucR no longer occludes GANTCs in target
379 promoters when cycling of CcrM is impaired (Fig 5C). Consistent with *SMa1635* and *SMa2245* being
380 MucR targets, LacZ-based promoter probe experiments (using P*a1635-lacZ* and P*a2245-lacZ*)
381 revealed that they are de-repressed in *S. meliloti mucR*::*Tn* cells compared to *WT* (Fig 5E) and that *S.*
382 *meliloti* MucR represses P*a1635-lacZ* and P*a2245-lacZ* in *C. crescentus WT* or Δ*mucR1/2* cells (Fig
383 5F). Importantly, when cycling of CcrM in *C. crescentus* was prevented by P_{lac}-*ccrM* or the *lon*
384 mutation P*a1635-lacZ* and P*a2245-lacZ* activity was increased compared to the *WT* strain (Fig 5G).
385 Thus, MucR controls hypomethylation during α-proteobacterial cell cycle.

386 **Fig 5. Hypomethylation control by MucR is conserved in α-proteobacteria.**

387 (A) Schematic of the two MucR-dependent hypomethylated loci (*SMa1635* and *SMa2245*) identified
388 by SMRT-sequencing in the in *S. meliloti WT* genome. Position of the hypomethylated GANTCs is
389 indicated by purple asterisks. The blue arrows indicate the DNA fragments cloned for LacZ promoter
390 probe assays. (B) *Hin*fI-qPCR analysis showing that *SMa1635* and *SMa2245* are hypomethylated in *S.*
391 *meliloti WT* cells compared to *mucR::Tn* cells. (C) Constitutive expression of $ccrM^{Cc}$ from P_{lac} on
392 pSRK [47] in *S. meliloti WT* cells increases the methylation percentage of *SMa1635* and *SMa2245*,
393 indicating that hypomethylation of GANTCs by MucR is also impaired in *S. meliloti* G1-phase cells.
394 (D) MucR occupancy at *SMa1635*, *SMa2245* and *SMc1552* (control) in *WT* and *mucR::Tn S. meliloti*
395 cells, as determined by qChIP using antibodies to *C. crescentus* MucR2. *SMa1635* and *SMa2245* are
396 bound by *S. meliloti* MucR, which suggests that hypomethylation of GANTCs at these loci is directly
397 due to occlusion by MucR. (E) Beta-galactosidase activity of P*a1635–lacZ* and P*a2245–lacZ* in *S.*
398 *meliloti* (fragments indicated by blue arrows in panel **A**). Both DNA fragments show a promoter
399 activity that is strongly de-repressed in *mucR::Tn* cells compared to the *WT* strain. Values are
400 expressed as percentages (activity in *WT* cells set at 100%). (F) Beta-galactosidase activity of P*a1635*

and P*a2245* in *C. crescentus WT* and Δ*mucR1/2* cells expressing $mucR^{Sm}$. Expression of $mucR^{Sm}$ from P_{van} on pMT335 decreases beta-galactosidase activity of P*a1635* and P*a2245*. Values are expressed as percentages (activity in *WT* cells carrying the empty vector set at 100%). (G) Beta-galactosidase activity of P*a1635* and P*a2245* in *C. crescentus WT,* P_{lac}*-ccrM* or *lon* cells. Values are expressed as percentages (activity in *WT* cells set at 100%).

Environmental and systemic signals controlling hypomethylation patterns.

To determine if other systemic (cell cycle) signals can alter methylation patterns in α-proteobacteria, we tested if CtrA can also occlude GANTC sites from methylation by CcrM. First, we constructed a synthetic promoter in which three GANTCs overlapping two CtrA-boxes (one GANTC in each CtrA box and one in between) were placed downstream of an attenuated *E. coli* phage T5 promoter on the *lacZ* promoter probe plasmid (Fig 6A). Next, we determined the methylation percentage of the GANTCs in *WT C. crescentus* cells harbouring the resulting reporter plasmid by *Hin*fI-qPCR analysis and found that the GANTCs are only partially methylated in *WT* cells, but efficiently methylated when CcrM is expressed ectopically (Fig 6A). Thus, methylation patterns can also emerge from competition between CcrM and other cell cycle regulators such as CtrA at appropriately positioned GANTCs.

Fig 6. Cell cycle and environmental signals affect methylation patterns.

(A) Competition between CtrA and CcrM. Schematic of the synthetic promoter carrying an attenuated *E. coli* phage T5 promoter followed by three GANTCs (purple asterisks) overlapping two CtrA-boxes (in yellow). *Hin*fI-qPCR analysis shows that this sequence is hypomethylated in *WT* cells, whereas constitutive expression of *ccrM* (P_{lac}*-ccrM*) increases the methylation percentage. This indicates that DNA-binding proteins other than MucR can also occlude GANTCs from methylation. (B) Methylation percentage of P*169* and P*1149* in phosphate-limiting conditions compared to rich medium (PYE), determined by *Hin*fI-qPCR analysis. Phosphate starvation (6h) significantly increases the methylation level of P*169* but not P*1149*. (C) Beta-galactosidase activity of P*169–lacZ* and *P169*–lacZ* (GANTCs

mutated to GTNTCs as in Fig 4A) in *WT* and Δ*mucR1/2* cells in rich medium and phosphate-limiting conditions. Phosphate starvation induces transcription from P*169–lacZ* and *P169*–lacZ* independently from the presence of MucR1/2. Values are expressed as percentages (activity in *WT* cells grown in PYE set at 100%). (D) Beta-galactosidase activity of P*1149–lacZ* in *WT* and Δ*mucR1/2* cells in rich medium and phosphate-limiting conditions. Phosphate starvation does not significantly affect the activity of P*1149–lacZ*. Values are expressed as percentages (activity in *WT* cells grown in PYE set at 100%).

To explore if environmental signals can also affect local hypomethylation patterns, we took advantage of the fact that expression of *CCNA_00169* (also known as *elpS*) is induced upon phosphate starvation of *C. crescentus* cells [44, 45]. Accordingly, we compared the P*169* methylation patterns by *Hin*fI-qPCR analysis of gDNA from *WT* cells grown in standard medium (PYE) and phosphate-limiting conditions. This revealed a significant increase in P*169* GANTC methylation in phosphate-limiting conditions compared to PYE (Fig 6B) and we observed a commensurate induction of P*169-lacZ* and P*169*-lacZ* that was MucR1/2 independent (Fig 6C). Both the increase in P*169* GANTC methylation and P*169-lacZ* activity are dependent on the phosphate-responsive transcriptional regulator PhoB (S4B-C Fig), suggesting that PhoB can facilitate methylation of MucR-protected GANTCs at P*169*. The result that no significant increase of the P*1149* methylation or P*1149-lacZ* activity was seen when *WT* cells were grown in phosphate-limiting conditions compared to standard PYE medium (or in Δ*phoB*::Ω cells compared to *WT*) argues against the possibility that changes in CcrM expression or activity underlie the modified methylation pattern of P*169* (Fig 6B, 6D; S4B-C Fig). Thus, P*169* provides an example of an environmental override for a promoter subject to local hypomethylation control by the cell cycle transcriptional circuitry.

Discussion

The correlative changes between human genetic variability and local (hypo)methylation prompt the question if and how such patterns are regulated by the cell cycle and/or environmental cues. Taking advantage of bacterial genomes that are small enough for full-methylome analysis by cutting-edge REC-Seq and SMRT-sequencing, we show that local m6A-hypomethylation exists in two different α-proteobacterial lineages and that conserved cell cycle factors govern its establishment in both systems. While in γ-proteobacterial lineages transcriptional regulators are also known to compete with the Dam m6A methylase to occlude certain methylation sites, local hypomethylation patterning has not been explored in the context of the transcriptional circuitry controlling progression of the (α-proteo)bacterial cell cycle. In eukaryotes methylation heterogeneity involves 5-methyl-cytosines introduced at CpG dinucleotides [2], but recently m6A marks, instated by unknown mechanisms, have also been reported [6-8]. Reliable detection of methylation sites by SMRT-sequencing requires extensive (25-fold) coverage for adenine methylation and even higher coverage for cytosine methylation (250-fold coverage needed in some instances)[5]. Non-methylated sites are only reliably detected by elimination of sites on which methylation is detected, thus leaving an element of uncertainty for those sites classified as non-methylated based on the absence of the kinetic signature for methylation. By contrast, REC-Seq with a methylation sensitive restriction enzyme was used here to enrich for non-methylated sites in α-proteobacteria by *Hin*fI cleavage. The continuum of scores we detected in these experiments points towards the use of REC-Seq in detecting loci whose methylation is variable within a culture, for example phase variable loci [48, 49]. *Hin*fI REC-Seq revealed the occurrence of non-methylated GANTCs in at least two α-proteobacterial genomes. Subsequent genetic analyses showed that the determinants controlling hypomethylation are conserved in these bacteria, but they are not encoded in eukaryotic genomes. However, at least one component, MucR, possesses an ancestral zinc-finger-type DNA binding domain [22], a protein domain which is also wide-spread in developmental regulation of eukaryotes [50]. The fact that MucR regulates expression of virulence and cell cycle genes [17-20], has recently been shown to confine genetic exchange by generalized transduction to G1-phase in *C. crescentus* via transcriptional regulation [41] and is responsible for hypomethylation of specific loci on the chromosome or megaplasmids thus raises the possibility that zinc-finger proteins may control (epigenetic) DNA transactions including local hypomethylation

480 during the eukaryotic cell cycle as well. In bacteria local methylation changes may correlate with
481 altered virulence behaviour and may underlie cell cycle-control in pathogens, endosymbionts or other
482 microbial systems. Methylation is known to influence virulence functions in γ-proteobacteria, often by
483 imposing bistability from phase-variable virulence promoters in subpopulations via transcriptional
484 regulators such as Lrp, Fur or OxyR [5, 9, 42, 48, 51-55]. Phenotypic heterogeneity in antibiotic drug
485 tolerance *in vivo* (a state known as persistence), which is acquired in a low fraction of bacterial cells,
486 may also underlie epigenetic changes induced stochastically by methylation, either deterministically
487 (during the cell cycle) or environmentally. Although no phase-variable promoters are currently known
488 for the α-proteobacteria, these bacteria offer the possibility to investigate the relationship of local
489 hypomethylation with cell cycle control, as both *C. crescentus* and *S. meliloti* are synchronizable and
490 exhibit comparable cell cycle control systems and transcription [13-15, 56]. However, as binding of
491 MucR to DNA is not impaired by methylation, the mechanisms underlying the increase in
492 transcription of target genes induced by methylation in α-proteobacteria (Fig 4D-F; Fig 5G) are likely
493 to be different from those described for γ-proteobacteria. Moreover, the α-proteobacteria lineage
494 includes the Rickettsiales order encompassing obligate intracellular pathogens, endosymbionts and the
495 extinct proto-mitochondrion from which the modern day mitochondria descended [57]. As MucR and
496 CcrM orthologs are not encoded in most Rickettsiales genomes, the determinants of hypomethylation
497 in this order must be different, if they do exist. Interestingly, endosymbionts from the genus
498 *Wolbachia* might provide a possible exception. Their genomes encode an unusual putative DNA
499 methylase in which a C-terminal pfam01555 methylase-domain is fused to a pfam02195 ParB-like
500 nuclease domain found in DNA-binding proteins and plasmid replication factors [58]. The sheltered
501 niche of obligate intracellular Rickettsia contrasts with that of free-living relatives that are exposed to
502 major environmental fluctuations.

503 In summary, our work shows that environmental regulatory responses like that to phosphate
504 limitation, which is particularly pertinent for bacteria living in aquatic ecosystems as *C. crescentus*,
505 are superimposed on (direct or indirect) hypo- or hyper-methylation control cued by the cell cycle. As
506 many hypomethylated sites occur upstream of genes encoding transcription factors (see S1 Table) and

transcription factors are often regulatory, it is conceivable that local hypomethylation is often induced by *cis*-encoded site-specific DNA-binding proteins that can compete with DNA methylases for overlapping target sites. The mechanism of DNA binding and temporal regulation of MucR remain to be elucidated in detail in order to reveal why MucR shields certain target sites from methylation by CcrM. Our work on MucR-dependent hypomethylation by *Hin*fI REC-Seq along with the comprehensive analysis of hypomethylated sites in other α-proteobacterial genomes [10] indicates that the functions controlled by hypomethylated promoters are distinct and generally not conserved among different α-proteobacteria. This suggests that hypomethylation does not play a major role in the regulation of the α-proteobacterial cell cycle, even though conserved cell cycle transcriptional regulators govern hypomethylation patterns. If it is largely serendipitous which sites MucR shields from methylation, it seems plausible that such hypomethylation control systems mediate species-specific transcriptional adaptations in response to stresses via MucR, CcrM or other variables that influence their binding, either directly or indirectly. For example, cell cycle controlled changes in local chromosomal topologies mediated by DNA replication or nucleoid-associated factors [59, 60] could exclude DNA methylases form specific target sites.

Materials and Methods

Strains and growth conditions

Caulobacter crescentus NA1000 [61] and derivatives were grown at 30°C in PYE (peptone-yeast extract) or M2G (minimal glucose). For phosphate starvation, *Caulobacter* cells were grown in 1/5X PYE (5-fold diluted PYE except 1 mM $MgSO_4$ and 1 mM $CaCl_2$, supplemented with 0.2% glucose). *Sinorhizobium meliloti* Rm2011 and derivatives were grown at 30°C in Luria broth (LB) supplemented with $CaCl_2$ 2.5 mM and $MgSO_4$ 2.5 mM. *Escherichia coli* S17-1 *λpir* and EC100D were grown at 37°C in LB. Swarmer cell isolation, electroporations, bi-parental matings and bacteriophage φCr30-mediated generalized transductions were performed as previously described [62-65]. Nalidixic acid, kanamycin, gentamicin and tetracycline were used at 20 (8 for *S. meliloti*), 20, 1 (10 for *E. coli*

533 and *S. meliloti*) and 1 (10 for *E. coli* and *S. meliloti*) μg/mL, respectively. Plasmids for β-galactosidase
534 assays were introduced into *S. meliloti* by bi-parental mating and into *C. crescentus* by electroporation.
535 Strains and plasmids constructions are detailed in the S1 Text file.

536

537 **Extraction of genomic DNA and methylation by qPCR (*Hin*fI-restriction/qPCR)**

538 Genomic DNA was extracted from mid-log phase cells (10 ml). Aliquots of DNA (0.5-1 μg)
539 were digested with *Hin*fI restriction endonuclease and used to determine the methylation percentage
540 by Real-Time PCR. Real-time PCR was performed using a Step-One Real-Time PCR system (Applied
541 Biosystems, Foster City, CA) using 0.05% of each DNA sample (5 μl of a dilution 1:100) digested
542 with *Hin*fI, 12.5 μl of SYBR green PCR master mix (Quanta Biosciences, Gaithersburg, MD) and
543 primers 10 μM each, in a total volume of 25 μl. A standard curve generated from the cycle threshold
544 (C_t) value of the serially diluted non-digested genomic DNA was used to calculate the methylation
545 percentage value for each sample. Average values are from triplicate measurements done per culture,
546 and the final data was generated from three independent cultures per strain and condition. The primers
547 used for Real-Time PCR are listed in Table B in the S1 Text file.

548

549 **Genome-wide methylation analyses**

550 SMRT (single-molecule real-time) sequencing libraries were prepared from gDNA extracted
551 from the four samples (*C. crescentus* and *S. meliloti WT* and *mucR* mutant strains) using the DNA
552 Template Prep Kit 2.0 (250bp - 3Kb, Pacific Biosciences p/n 001-540-726). Sequences generated by
553 the Pacific Bioscience RSII were aligned to the *C. crescentus* NA1000 or *S. meliloti* Rm2011 genomes
554 [37, 66, 67] using Blasr (https://github.com/PacificBiosciences/blasr) and the modification and
555 associated motifs patterns were identified applying the RS_Modification_and_Motif_Analyisis
556 protocol in SMRT Analysis (https://github.com/PacificBiosciences/SMRT-Analysis/wiki/SMRT-
557 Analysis-Software-Installation-v2.2.0). For each aligned base, a statistics measured as interpulse

duration (IPD) combined with a modification quality value (QV) would mark the methylation status. On the one hand, a minimum QV of 45 is required for a position to be marked as methylated; on the other hand, a maximum QV between 10 and 30 (depending on the observed kinetic detections background in the sample), coupled with the requirement that such a score is observed on both strands, would mean that a position, in an otherwise methylated motif, is unmethylated.

For REC-Seq (*r*estriction *e*nzyme *c*leavage–*seq*uencing) 1 µg of genomic DNA from *C. crescentus* NA1000 and *S. meliloti* Rm2011 was cleaved with *Hin*fI, a blocked (5'biotinylated) specific adaptor was ligated to the ends and the ligated fragments were then sheared to an average size of 150-400 bp (Fasteris SA, Geneva, CH). Illumina adaptors were then ligated to the sheared ends followed by deep-sequencing using a Hi-Seq Illumina sequencer, and the (50 bp single end) reads were quality controlled with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). To remove contaminating sequences, the reads were split according to the *Hin*fI consensus motif (5'-G^ANTC-3') considered as a barcode sequence using fastx_toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) (fastx_barcode_splitter.pl --bcfile barcodelist.txt --bol --exact). Most of the reads (more than 90%) were rejected, and the reads kept were remapped to the reference genomes [37, 66, 67] with bwa mem [68] and samtools [69] to generate a sorted bam file. The bam file was further filtered to remove low mapping quality reads (keeping AS >= 45) and split by orientation (alignmentFlag 0 or 16) with bamtools [70]. The reads were counted at 5' positions using Bedtools [71] (bedtools genomecov -d -5). Both orientation count files were combined into a bed file at each identified 5'-GANTC-3' motif (where reverse counts >=1 at position N+1 and forward counts >=1 at position N-1) using a home-made PERL script. The *Hin*fI positions in the bed file were associated with the closest gene using Bedtools closest [71] and the gff3 file of the reference genomes [72]. The final bed file was converted to an MS Excel sheet (S1 and S2 Tables) with a homemade script. For the *Mbo*I-based REC-Seq, the strategy was identical except that a different adaptor was used for ligation after cleavage and the *Mbo*I consensus motif (5'-^GATC-3') was used as barcode for filtering of *V. cholerae* O1 biovar El Tor [73] and *E. coli* K12 Ec100D gDNA mapped onto the MG1655 genome [74].

585
586 **ChIP-Exo on *C. crescentus* synchronized cells**

587 *C. crescentus WT* cells for ChIP-Exo were taken at different time points after synchronization
588 (10, 40, 70 and 100 minutes). After cross-linking, chromatin was prepared as previously described
589 [17]. ChIP-Exo was performed with 2 µl of polyclonal antibodies to MucR1 at Peconic LCC
590 (http://www.peconicgenomics.com) (State College, PA), which provided standard genomic position
591 format files (BAM) as output using the SOLiD genome sequencer (Applied Biosystems). A custom
592 Perl script was then used to calculate the sequencing (read) coverage per base (*per-base* coverage) for
593 each ChIP-Exo sample. Next, we computed the enrichment ratio (ER) for each promoter region. To
594 this end, the Perl script extracted the *per-base* coverage of a 600 bp region for each ORF (from -500 to
595 +100 from the start codon for each ORF annotated in *C. crescentus* genome) and calculated the
596 average coverage for each of these regions. The resulting value was then normalized with respect to
597 the coverage of all the intergenic regions. This was done (by the Perl script) by selecting all the
598 intergenic regions in the *C. crescentus* genome, merging them and extracting the *per-base* coverage
599 for all these intergenic regions. The coverage was averaged for windows of 600 bp, shifting each
600 window by 100 bp, and the mean of all resulting values was computed. The ER for each promoter
601 region was therefore calculated as the ratio between the average coverage of the promoter region
602 divided by the mean obtained for the intergenic regions.

603

604 **Transcriptional start sites mapping by exact mapping of transcriptome ends (EMOTE)**

605 The transcription start sites in the NA1000 *WT* and the Δ*mucR1/2* mutant were determined by
606 TSS-EMOTE (Transcription Start Specific Exact Mapping Of Transcriptome Ends), a global assay
607 that reveals the sequence of the 20 first nucleotides of 5'-triphosphorylated RNA in a sample based on
608 an XRN-1 digest of transcripts lacking the 5' triphosphate ends [46]. The TSS-EMOTE protocol and
609 analyses were performed according to the scheme in S3 Fig and the detailed protocol described in
610 [75]. We used a Worst-Case (i.e.) smallest difference) model to compare the number of Unique
611 Molecular Identifiers between the two pairs of biological replicates (i.e. mutant vs. wild-type) and
612 provide additional information about relative expression for each of the detected TSSs. The full list of

613 detected TSSs is shown in S5 Table and TSSs at the relevant genomic loci are indicated by black
614 (sense) and green (antisense) arrows in Fig 2A.
615

616 ## β-galactosidase assays

617 β-galactosidase assays were performed at 30°C. Cells (50-200 µl) at OD_{660nm}=0.1-0.5 were
618 lysed with chloroform and mixed with Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl and
619 1 mM $MgSO_4$, pH 7) to a final volume of 800 µl. Two hundred µl of ONPG (o-nitrophenyl-β-D-
620 galactopyranoside, stock solution 4 mg/ml in 0.1 M potassium phosphate, pH 7) were added and the
621 reaction timed. When a medium-yellow colour developed, the reaction was stopped by adding 400 µl
622 of 1M Na_2CO_3. The OD_{420nm} of the supernatant was determined and the Miller units (U) were
623 calculated as follows: U= (OD_{420nm}*1000)/(OD_{660nm}*time [in min] *volume of culture used [in ml]).
624 Error was computed as standard deviation (SD) of at least three independent experiments.
625

626 ## qChIP assay on *S. meliloti*

627 Samples for qChIP assay were prepared from mid-log phase *S. meliloti* cells as previously
628 described [17]. Two microliters of polyclonal antibodies to MucR2 were used for the
629 immunoprecipitation.

630 Real-time PCR was performed as described for *Hinf*I-restricted genomic DNA, using 0.5% of
631 each ChIP sample (5 µl of a dilution 1:10). A standard curve generated from the cycle threshold (C_t)
632 value of the serially diluted chromatin input was used to calculate the percentage input value for each
633 sample. Average values are from triplicate measurements done per culture, and the final data was
634 generated from three independent cultures per strain. The primers used for *SMa1635* and *SMa2245*
635 loci were the same as for the determination of the methylation percentage of these loci (Table B in S1
636 Text file).
637

638 ## Immunoblots

639 For immunoblots, protein samples were separated on SDS polyacrylamide gel, transferred to
640 polyvinylidene difluoride (PVDF) Immobilon-P membranes (Merck Millipore) and blocked in PBS
641 (phosphate saline buffer) 0.1% Tween20 and 5% dry milk. The anti-sera were used at the following
642 dilutions: anti-CcrM (1:10'000) [26], anti-MucR1 [17] (1:10'000), anti-MucR2 [17] (1:10'000).
643 Protein-primary antibody complexes were visualized using horseradish peroxidase-labelled anti-rabbit
644 antibodies and ECL detection reagents (Merck Millipore).

645

646 Plasmids, primers, synthetic fragments and strains constructions are described in the S1 Text file.

647

648 **Data Access**

649 Deep-sequencing data are deposited in Gene Expression Omnibus database (GEO: GSE79880).

650

651

652 ## Acknowledgments

653 We thank Julien Prados for help with TSS-EMOTE bioinformatics and Laurence Théraulaz
654 for excellent technical assistance. We thank Melanie Blokesch and Yoshiharu Yamaichi for providing
655 *V. cholerae* gDNA.

656

References

1. Casadesus J, Low D. Epigenetic gene regulation in the bacterial world. Microbiol Mol Biol Rev. 2006;70(3):830-56. doi: 10.1128/MMBR.00016-06. PubMed PMID: 16959970.
2. Schubeler D. Function and information content of DNA methylation. Nature. 2015;517(7534):321-6. doi: 10.1038/nature14192. PubMed PMID: 25592537.
3. Arber W, Dussoix D. Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. J Mol Biol. 1962;5:18-36. PubMed PMID: 13862047.
4. Marinus MG, Casadesus J. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. FEMS Microbiol Rev. 2009;33(3):488-503. doi: 10.1111/j.1574-6976.2008.00159.x. PubMed PMID: 19175412.
5. Sanchez-Romero MA, Cota I, Casadesus J. DNA methylation in bacteria: from the methyl group to the methylome. Curr Opin Microbiol. 2015;25:9-16. doi: 10.1016/j.mib.2015.03.004. PubMed PMID: 25818841.
6. Greer EL, Blanco MA, Gu L, Sendinc E, Liu J, Aristizabal-Corrales D, et al. DNA Methylation on N6-Adenine in *C. elegans*. Cell. 2015;161(4):868-78. doi: 10.1016/j.cell.2015.04.005. PubMed PMID: 25936839.
7. Fu Y, Luo GZ, Chen K, Deng X, Yu M, Han D, et al. N6-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. Cell. 2015;161(4):879-92. doi: 10.1016/j.cell.2015.04.010. PubMed PMID: 25936837.
8. Zhang G, Huang H, Liu D, Cheng Y, Liu X, Zhang W, et al. N6-methyladenine DNA modification in Drosophila. Cell. 2015;161(4):893-906. doi: 10.1016/j.cell.2015.04.018. PubMed PMID: 25936838.
9. Adhikari S, Curtis PD. DNA methyltransferases and epigenetic regulation in bacteria. FEMS Microbiol Rev. 2016;40(5):575-91. doi: 10.1093/femsre/fuw023. PubMed PMID: 27476077.
10. Blow MJ, Clark TA, Daum CG, Deutschbauer AM, Fomenkov A, Fries R, et al. The Epigenomic Landscape of Prokaryotes. PLoS Genet. 2016;12(2):e1005854. doi: 10.1371/journal.pgen.1005854. PubMed PMID: 26870957.
11. Collier J. Epigenetic regulation of the bacterial cell cycle. Curr Opin Microbiol. 2009;12(6):722-9. doi: 10.1016/j.mib.2009.08.005. PubMed PMID: 19783470.
12. Kozdon JB, Melfi MD, Luong K, Clark TA, Boitano M, Wang S, et al. Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle. Proc Natl Acad Sci U S A. 2013;110(48):E4658-67. doi: 10.1073/pnas.1319315110. PubMed PMID: 24218615.
13. Panis G, Murray SR, Viollier PH. Versatility of global transcriptional regulators in alpha-Proteobacteria: from essential cell cycle control to ancillary functions. FEMS Microbiol Rev. 2015;39(1):120-33. doi: 10.1093/femsre/fuu002. PubMed PMID: 25793963.
14. De Nisco NJ, Abo RP, Wu CM, Penterman J, Walker GC. Global analysis of cell cycle gene expression of the legume symbiont *Sinorhizobium meliloti*. Proc Natl Acad Sci U S A. 2014;111(9):3217-24. doi: 10.1073/pnas.1400421111. PubMed PMID: 24501121.
15. Laub MT, McAdams HH, Feldblyum T, Fraser CM, Shapiro L. Global analysis of the genetic network controlling a bacterial cell cycle. Science. 2000;290(5499):2144-8. doi: 8994 [pii]. PubMed PMID: 11118148.
16. Hallez R, Bellefontaine AF, Letesson JJ, De Bolle X. Morphological and functional asymmetry in alpha-proteobacteria. Trends Microbiol. 2004;12(8):361-5. doi: 10.1016/j.tim.2004.06.002. PubMed PMID: 15276611.
17. Fumeaux C, Radhakrishnan SK, Ardissone S, Theraulaz L, Frandi A, Martins D, et al. Cell cycle transition from S-phase to G1 in *Caulobacter* is mediated by ancestral virulence regulators. Nature communications. 2014;5:4081. doi: 10.1038/ncomms5081. PubMed PMID: 24939058.
18. Mirabella A, Terwagne M, Zygmunt MS, Cloeckaert A, De Bolle X, Letesson JJ. *Brucella melitensis* MucR, an Orthologue of *Sinorhizobium meliloti* MucR, Is Involved in Resistance to Oxidative, Detergent, and Saline Stresses and Cell Envelope Modifications. J Bacteriol. 2013;195(3):453-65. doi: 10.1128/JB.01336-12. PubMed PMID: 23161025.

19. Cooley MB, Kado CI. Mapping of the ros virulence regulatory gene of *A. tumefaciens*. Mol Gen Genet. 1991;230(1-2):24-7. PubMed PMID: 1660566.
20. Mueller K, Gonzalez JE. Complex regulation of symbiotic functions is coordinated by MucR and quorum sensing in *Sinorhizobium meliloti*. J Bacteriol. 2011;193(2):485-96. doi: 10.1128/JB.01129-10. PubMed PMID: 21057009.
21. Baglivo I, Russo L, Esposito S, Malgieri G, Renda M, Salluzzo A, et al. The structural role of the zinc ion can be dispensable in prokaryotic zinc-finger domains. Proc Natl Acad Sci U S A. 2009;106(17):6933-8. Epub 2009/04/17. doi: 10.1073/pnas.0810003106. PubMed PMID: 19369210.
22. Malgieri G, Russo L, Esposito S, Baglivo I, Zaccaro L, Pedone EM, et al. The prokaryotic Cys2His2 zinc-finger adopts a novel fold as revealed by the NMR structure of *Agrobacterium tumefaciens* Ros DNA-binding domain. Proc Natl Acad Sci U S A. 2007;104(44):17341-6. doi: 10.1073/pnas.0706659104. PubMed PMID: 17956987.
23. Domian IJ, Quon KC, Shapiro L. Cell type-specific phosphorylation and proteolysis of a transcriptional regulator controls the G1-to-S transition in a bacterial cell cycle. Cell. 1997;90(3):415-24.
24. Quon KC, Marczynski GT, Shapiro L. Cell cycle control by an essential bacterial two-component signal transduction protein. Cell. 1996;84(1):83-93.
25. Reisenauer A, Quon K, Shapiro L. The CtrA response regulator mediates temporal control of gene expression during the Caulobacter cell cycle. J Bacteriol. 1999;181(8):2430-9.
26. Stephens C, Reisenauer A, Wright R, Shapiro L. A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. Proc Natl Acad Sci USA. 1996;93(3):1210-4.
27. Laub MT, Chen SL, Shapiro L, McAdams HH. Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. Proc Natl Acad Sci U S A. 2002;99(7):4632-7. Epub 2002/04/04. doi: 10.1073/pnas.062065699. PubMed PMID: 11930012.
28. Wright R, Stephens C, Zweiger G, Shapiro L, Alley MR. *Caulobacter* Lon protease has a critical role in cell-cycle control of DNA methylation. Genes Dev. 1996;10(12):1532-42.
29. Jonas K, Liu J, Chien P, Laub MT. Proteotoxic stress induces a cell-cycle arrest by stimulating Lon to degrade the replication initiator DnaA. Cell. 2013;154(3):623-36. doi: 10.1016/j.cell.2013.06.034. PubMed PMID: 23911325.
30. Collier J, McAdams HH, Shapiro L. A DNA methylation ratchet governs progression through a bacterial cell cycle. Proc Natl Acad Sci U S A. 2007;104(43):17111-6. doi: 10.1073/pnas.0708112104. PubMed PMID: 17942674.
31. Lluch-Senar M, Luong K, Llorens-Rico V, Delgado J, Fang G, Spittle K, et al. Comprehensive methylome characterization of *Mycoplasma genitalium* and *Mycoplasma pneumoniae* at single-base resolution. PLoS Genet. 2013;9(1):e1003191. doi: 10.1371/journal.pgen.1003191. PubMed PMID: 23300489.
32. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, et al. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. Nat Biotechnol. 2012;30(12):1232-9. doi: 10.1038/nbt.2432. PubMed PMID: 23138224.
33. Zweiger G, Marczynski G, Shapiro L. A *Caulobacter* DNA methyltransferase that functions only in the predivisional cell. J Mol Biol. 1994;235(2):472-85.
34. Dalia AB, Lazinski DW, Camilli A. Characterization of undermethylated sites in *Vibrio cholerae.* J Bacteriol. 2013;195(10):2389-99. doi: 10.1128/JB.02112-12. PubMed PMID: 23504020.
35. Kahng LS, Shapiro L. The CcrM DNA methyltransferase of *Agrobacterium tumefaciens* is essential, and its activity is cell cycle regulated. J Bacteriol. 2001;183(10):3065-75. Epub 2001/04/28. doi: 10.1128/JB.183.10.3065-3075.2001. PubMed PMID: 11325934.
36. Robertson GT, Reisenauer A, Wright R, Jensen RB, Jensen A, Shapiro L, et al. The *Brucella abortus* CcrM DNA methyltransferase is essential for viability, and its overexpression attenuates intracellular replication in murine macrophages. J Bacteriol. 2000;182(12):3482-9.
37. Capela D, Barloy-Hubler F, Gouzy J, Bothe G, Ampe F, Batut J, et al. Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. Proc Natl Acad Sci U S A. 2001;98(17):9877-82. doi: 10.1073/pnas.161294398. PubMed PMID: 11481430.

762 38. Thanbichler M, Iniesta AA, Shapiro L. A comprehensive set of plasmids for vanillate- and
763 xylose-inducible gene expression in *Caulobacter crescentus*. Nucleic Acids Research.
764 2007;35(20):e137. doi: 10.1093/nar/gkm818.
765 39. Mahony S, Pugh BF. Protein-DNA binding in high-resolution. Crit Rev Biochem Mol Biol.
766 2015;50(4):269-83. doi: 10.3109/10409238.2015.1051505. PubMed PMID: 26038153.
767 40. Fioravanti A, Fumeaux C, Mohapatra SS, Bompard C, Brilli M, Frandi A, et al. DNA binding
768 of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in *Caulobacter*
769 *crescentus* and other Alphaproteobacteria. PLoS Genet. 2013;9(5):e1003541. doi:
770 10.1371/journal.pgen.1003541. PubMed PMID: 23737758.
771 41. Ardissone S, Fumeaux C, Bergé M, Beaussart A, Theraulaz L, Radhakrishnan SK, et al. Cell
772 cycle constraints on capsulation and bacteriophage susceptibility. eLife. 2014;3. doi:
773 10.7554/eLife.03587. PubMed PMID: 25421297.
774 42. Cota I, Bunk B, Sproer C, Overmann J, Konig C, Casadesus J. OxyR-dependent formation of
775 DNA methylation patterns in OpvABOFF and OpvABON cell lineages of *Salmonella enterica*.
776 Nucleic Acids Res. 2015. doi: 10.1093/nar/gkv1483. PubMed PMID: 26687718.
777 43. Gonzalez D, Kozdon JB, McAdams HH, Shapiro L, Collier J. The functions of DNA
778 methylation by CcrM in *Caulobacter crescentus*: a global approach. Nucleic Acids Res. 2014. doi:
779 10.1093/nar/gkt1352. PubMed PMID: 24398711.
780 44. Le Blastier S, Hamels A, Cabeen M, Schille L, Tilquin F, Dieu M, et al. Phosphate starvation
781 triggers production and secretion of an extracellular lipoprotein in *Caulobacter crescentus*. PLoS One.
782 2010;5(12):e14198. doi: 10.1371/journal.pone.0014198. PubMed PMID: 21152032.
783 45. Lubin EA, Henry JT, Fiebig A, Crosson S, Laub MT. Identification of the PhoB Regulon and
784 Role of PhoU in the Phosphate Starvation Response of *Caulobacter crescentus*. J Bacteriol.
785 2015;198(1):187-200. doi: 10.1128/JB.00658-15. PubMed PMID: 26483520.
786 46. Redder P. Using EMOTE to map the exact 5'-ends of processed RNA on a transcriptome-wide
787 scale. Methods Mol Biol. 2015;1259:69-85. doi: 10.1007/978-1-4939-2214-7_5. PubMed PMID:
788 25579580.
789 47. Khan SR, Gaines J, Roop RM, 2nd, Farrand SK. Broad-host-range expression vectors with
790 tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti
791 plasmid quorum sensing. Appl Environ Microbiol. 2008;74(16):5053-62. doi: 10.1128/AEM.01098-
792 08. PubMed PMID: 18606801.
793 48. Cota I, Blanc-Potard AB, Casadesus J. STM2209-STM2208 (opvAB): a phase variation locus
794 of *Salmonella enterica* involved in control of O-antigen chain length. PLoS One. 2012;7(5):e36863.
795 doi: 10.1371/journal.pone.0036863. PubMed PMID: 22606300.
796 49. Casadesus J, Low DA. Programmed heterogeneity: epigenetic mechanisms in bacteria. J Biol
797 Chem. 2013;288(20):13929-35. doi: 10.1074/jbc.R113.472274. PubMed PMID: 23592777.
798 50. Klug A. The discovery of zinc fingers and their applications in gene regulation and genome
799 manipulation. Annual review of biochemistry. 2010;79:213-31. doi: 10.1146/annurev-biochem-
800 010909-095056. PubMed PMID: 20192761.
801 51. Cota I, Sanchez-Romero MA, Hernandez SB, Pucciarelli MG, Garcia-Del Portillo F,
802 Casadesus J. Epigenetic Control of *Salmonella enterica* O-Antigen Chain Length: A Tradeoff between
803 Virulence and Bacteriophage Resistance. PLoS Genet. 2015;11(11):e1005667. doi:
804 10.1371/journal.pgen.1005667. PubMed PMID: 26583926.
805 52. Arnoldini M, Vizcarra IA, Pena-Miller R, Stocker N, Diard M, Vogel V, et al. Bistable
806 expression of virulence genes in salmonella leads to the formation of an antibiotic-tolerant
807 subpopulation. PLoS Biol. 2014;12(8):e1001928. doi: 10.1371/journal.pbio.1001928. PubMed PMID:
808 25136970.
809 53. Broadbent SE, Davies MR, van der Woude MW. Phase variation controls expression of
810 *Salmonella* lipopolysaccharide modification genes by a DNA methylation-dependent mechanism. Mol
811 Microbiol. 2010;77(2):337-53. doi: 10.1111/j.1365-2958.2010.07203.x. PubMed PMID: 20487280.
812 54. Peterson SN, Reich NO. Competitive Lrp and Dam assembly at the pap regulatory region:
813 implications for mechanisms of epigenetic regulation. J Mol Biol. 2008;383(1):92-105. doi:
814 10.1016/j.jmb.2008.07.086. PubMed PMID: 18706913.
815 55. Atack JM, Srikhanta YN, Fox KL, Jurcisek JA, Brockman KL, Clark TA, et al. A biphasic
816 epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable

817 *Haemophilus influenzae*. Nature communications. 2015;6:7828. doi: 10.1038/ncomms8828. PubMed
818 PMID: 26215614.
819 56. Pini F, De Nisco NJ, Ferri L, Penterman J, Fioravanti A, Brilli M, et al. Cell Cycle Control by
820 the Master Regulator CtrA in *Sinorhizobium meliloti*. PLoS Genet. 2015;11(5):e1005232. doi:
821 10.1371/journal.pgen.1005232. PubMed PMID: 25978424.
822 57. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski
823 RM, et al. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature.
824 1998;396(6707):133-40. doi: 10.1038/24094. PubMed PMID: 9823893.
825 58. Ellegaard KM, Klasson L, Naslund K, Bourtzis K, Andersson SG. Comparative genomics of
826 Wolbachia and the bacterial species concept. PLoS Genet. 2013;9(4):e1003381. doi:
827 10.1371/journal.pgen.1003381. PubMed PMID: 23593012.
828 59. Rimsky S, Travers A. Pervasive regulation of nucleoid structure and function by nucleoid-
829 associated proteins. Curr Opin Microbiol. 2011;14(2):136-41. doi: 10.1016/j.mib.2011.01.003.
830 PubMed PMID: 21288763.
831 60. Sobetzko P, Travers A, Muskhelishvili G. Gene order and chromosome dynamics coordinate
832 spatiotemporal gene expression during the bacterial growth cycle. Proceedings of the National
833 Academy of Sciences. 2012;109(2):E42–E50. doi: 10.1073/pnas.1108229109.
834 61. Evinger M, Agabian N. Envelope-associated nucleoid from *Caulobacter crescentus* stalked
835 and swarmer cells. J Bacteriol. 1977;132(1):294-301.
836 62. Ely B. Genetics of *Caulobacter crescentus*. Methods Enzymol. 1991;204:372-84.
837 63. Chen JC, Viollier PH, Shapiro L. A membrane metalloprotease participates in the sequential
838 degradation of a *Caulobacter* polarity determinant. Mol Microbiol. 2005;55(4):1085-103. Epub
839 2005/02/03. doi: 10.1111/j.1365-2958.2004.04443.x. PubMed PMID: 15686556.
840 64. Viollier PH, Shapiro L. A lytic transglycosylase homologue, PleA, is required for the
841 assembly of pili and the flagellum at the *Caulobacter crescentus* cell pole. Mol Microbiol.
842 2003;49(2):331-45. Epub 2003/06/28. doi: 3576 [pii]. PubMed PMID: 12828633.
843 65. Viollier PH, Thanbichler M, McGrath PT, West L, Meewan M, McAdams HH, et al. Rapid
844 and sequential movement of individual chromosomal loci to specific subcellular locations during
845 bacterial DNA replication. Proc Natl Acad Sci U S A. 2004;101(25):9257-62. doi:
846 10.1073/pnas.0402606101. PubMed PMID: 15178755.
847 66. Marks ME, Castro-Rojas CM, Teiling C, Du L, Kapatral V, Walunas TL, et al. The genetic
848 basis of laboratory adaptation in *Caulobacter crescentus*. J Bacteriol. 2010;192(14):3678-88. doi:
849 10.1128/JB.00255-10. PubMed PMID: 20472802.
850 67. Nierman WC, Feldblyum TV, Laub MT, Paulsen IT, Nelson KE, Eisen J, et al. Complete
851 genome sequence of *Caulobacter crescentus*. Proc Natl Acad Sci U S A. 2001;98(7):4136-41.
852 68. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform.
853 Bioinformatics. 2010;26(5):589-95. doi: 10.1093/bioinformatics/btp698. PubMed PMID: 20080505;
854 PubMed Central PMCID.
855 69. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and
856 population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27(21):2987-
857 93. doi: 10.1093/bioinformatics/btr509. PubMed PMID: 21903627.
858 70. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. BamTools: a C++ API
859 and toolkit for analyzing and managing BAM files. Bioinformatics. 2011;27(12):1691-2. doi:
860 10.1093/bioinformatics/btr174. PubMed PMID: 21493652.
861 71. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
862 Bioinformatics. 2010;26(6):841-2. doi: 10.1093/bioinformatics/btq033. PubMed PMID: 20110278.
863 72. Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, et al. Ensembl Genomes
864 2016: more genomes, more complexity. Nucleic Acids Res. 2016;44(D1):D574-80. doi:
865 10.1093/nar/gkv1209. PubMed PMID: 26578574.
866 73. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, et al. DNA
867 sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature. 2000;406(6795):477-
868 83.
869 74. Riley M, Abe T, Arnaud MB, Berlyn MKB, Blattner FR, Chaudhuri RR, et al. *Escherichia*
870 *coli* K-12: a cooperatively developed annotation snapshot—2005. Nucleic Acids Research.
871 2006;34(1):1-9. doi: 10.1093/nar/gkj405.

872 75. Prados J., Linder P., Redder P. TSS-EMOTE, a refined protocol for a more complete and less
873 biased global mapping of trnscription start sites in bacterial pathogens. BMC genomics. 2016;17:849.
874 doi: 10.1186/s12864-016-3211-3.

875

876 **S1 Text file**

877 **Strains and plasmids construction.**

878 Table A. Strains and plasmids used in this study.

879 Table B. Oligonucleotides used in this study.

880

881 **S1 Fig. Controls for methylation percentage by *Hin*fI-qPCR and MucR-dependent**
882 **antisense transcription.**

883 (A) Methylation percentage of control loci in the *WT* and Δ*mucR1/2* strains, as determined by *Hin*fI-
884 qPCR: P_{nagA} (hypomethylated, MucR-independent), P_{xylX} (no GANTCs), $P_{pod}J$ (fully methylated,
885 MucR-independent). The controls show that MucR does not affect the methylation of sequences that
886 are not its direct targets. (Note that as P_{xylX} contains no GANTCs, the label on the y-axis should not be
887 interpreted as methylation but cleavage percentage). (B) Methylation percentage of P*1248* and P*3426*
888 determined by *Hin*fI-qPCR analysis. The graphs show that these sequences are hypomethylated in the
889 Δ*mucR1/2* compared to the *WT* strain, as predicted by the SMRT-sequencing and REC-Seq. This
890 suggests that other DNA-binding proteins, directly or indirectly MucR-dependent, can also occlude
891 GANTCs from methylation. (C) Immunoblot anti-CcrM in *WT* and *lon* mutant cells carrying an empty
892 vector, P_{van}-*mucR1* or P_{van}-*mucR1* long (dominant negative form of MucR1 with an N-terminal
893 extension). Over-expression of MucR1 does not affect the steady state levels of CcrM (arrow). Note
894 that CcrM levels are elevated in the *lon* mutant strain as the protein is stabilized. Molecular size
895 standards are indicated on the right as blue lines with the corresponding values in kDa. (D)
896 Methylation percentage of the *parS* locus in *C. crescentus WT* cells carrying the empty vector, the
897 heterologous methylases (*TA*, *HP*, *hinf*) under control of P_{van} on pMT335, P_{lac}-*ccrM* and in the *lon*Ω
898 mutant. Stabilisation of CcrM or constitutive expression of heterologous methylases does not affect
899 the methylation state of GANTCs at the *parS* locus. (E) Beta-galactosidase activity of two MucR-

900 dependent antisense promoters identified by TSS-EMOTE. Values are expressed as percentages
901 (activity in *WT* cells set at 100%).

902

903 **S2 Fig. Rec-Seq comparison of GANTC methylation in *WT* and Δ*mucR1/2*.**

904 The Rec-Seq score of each GANTC site in *WT* and Δ*mucR1/2 C. crescentus* was normalized according
905 to the total number of reads obtained for the strains. The graph represents the difference between the
906 normalized score obtained for the *WT* and the normalized score for the Δ*mucR1/2* strain for each
907 GANTC site according to the position along the chromosome. Positive values (in blue) indicate
908 hypomethylation in the *WT*, whereas negative values (in red) indicate hypomethylation in the
909 Δ*mucR1/2* strain. The GANTCs verified by *Hin*fI-qPCR are indicated.

910

911 **S3 Fig. Flowchart of the EMOTE assay.**

912 RNA is shown as thin lines and DNA as thick lines, double lines represent Illumina adaptors. (A)
913 Cellular RNAs exist as primary transcripts with a triphosphorylated 5'end (PPP, red line) and
914 processed transcripts with either a monophosphorylated 5'end (P, dashed grey line) or a non-
915 phosphorylated 5'-OH end (straight grey line). The asterisk indicates the ends of interest. XRN1 is
916 used to remove 5' monophosphorylated RNA (grey dotted line) from the total RNA samples. (B)
917 Treatment with *E. coli* RppH converts the 5' triphosphorylated end of primary transcripts to a
918 monophosphorylated 5'end, a substrate for ligation (C, D) to the Rp6 synthetic oligonucleotide with
919 T4 RNA ligase 1, which does not accept triphosphorylated or non-phosphorylated substrates. A mock
920 reaction is performed at this stage in the absence of RppH to control for background (non-specific)
921 signals. (E, F) Reverse transcription generates cDNA from both non-phosphorylated and Rp6-ligated
922 RNA. Open arrows indicate polymerase extension. (G) Only cDNA from Rp6-ligated (and therefore
923 originally triphosphorylated) RNA is amplified by the primers that add EMOTE barcodes (xxx) and
924 Illumina adaptors. (H) Illumina sequencing (50 nucleotides) from the "A" end (see panel G**)** results in

925 reads that have the specific EMOTE barcode of the original RNA sample, the Rp6 sequence and the
926 first 20 nucleotides of the original triphosphorylated RNA, permitting exact identification of the
927 original 5' end.

928

929 **S4 Fig. Antibodies against *C. crescentus* MucR2 specifically recognize *S. meliloti* MucR.**
930 **Methylation and induction of P*169* are PhoB-dependent.**

931 (A) Immunoblot on *C. crescentus* and *S. meliloti* total cell extracts showing that the polyclonal
932 antibodies against *C. crescentus* MucR2 specifically recognize *S. meliloti* MucR (SMc00058).
933 Molecular size standards are indicated on the right as blue lines with the corresponding values in kDa.
934 (B) Methylation percentage of P*169* and P*1149* in phosphate-limiting conditions compared to rich
935 medium (PYE) in *WT* and Δ*phoB*::Ω cells, determined by *Hinf*I-qPCR analysis. The graphs show that
936 the increase in the methylation state of P*169* is dependent on the presence of the conserved
937 transcriptional regulator PhoB. (C) Beta-galactosidase activity of P*169–lacZ* and P*1149–lacZ* in *WT*
938 and Δ*phoB*:: Ω cells in rich medium and phosphate-limiting conditions. The induction of P*169-lacZ* is
939 specifically dependent on the presence of PhoB. Values are expressed as percentages (activity in *WT*
940 cells grown in PYE set at 100%).

941

942 **S1 Table.** REC-Seq *Hin*fI analysis of *C. crescentus* and *S. meliloti* genomic DNA. In both cases, wild
943 type and *mucR* mutant strains were analysed. For the GANTC sites with a score higher than 100 in the
944 *WT C. crescentus*, REC-Seq data were compared to available ChIP-Seq [17] and SMRT-Sequencing
945 data [12]. The GANTCs tested by HinfI-qPCR assay are highlighted in yellow. The 50 GANTCs with
946 the highest score in *WT C. crescentus* are those shown also in Table 1.

947 **S2 Table.** REC-Seq *Mbo*I analysis of *E. coli K12* and *Vibrio cholerae* genomic DNA.

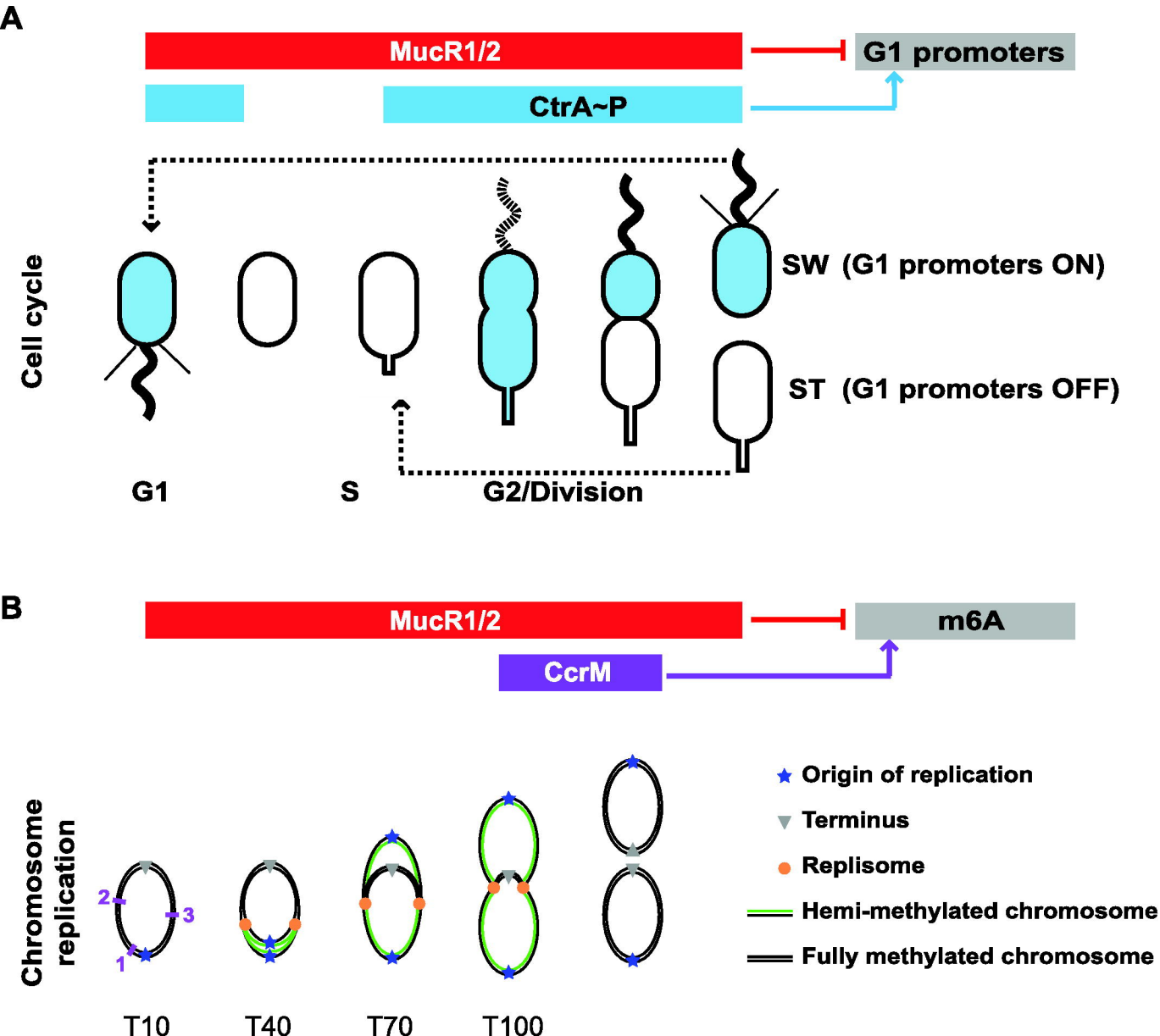
948 **S3 Table.** Non-methylated GANTCs predicted from SMRT analysis of *WT* and *mucR* mutant in *C.*

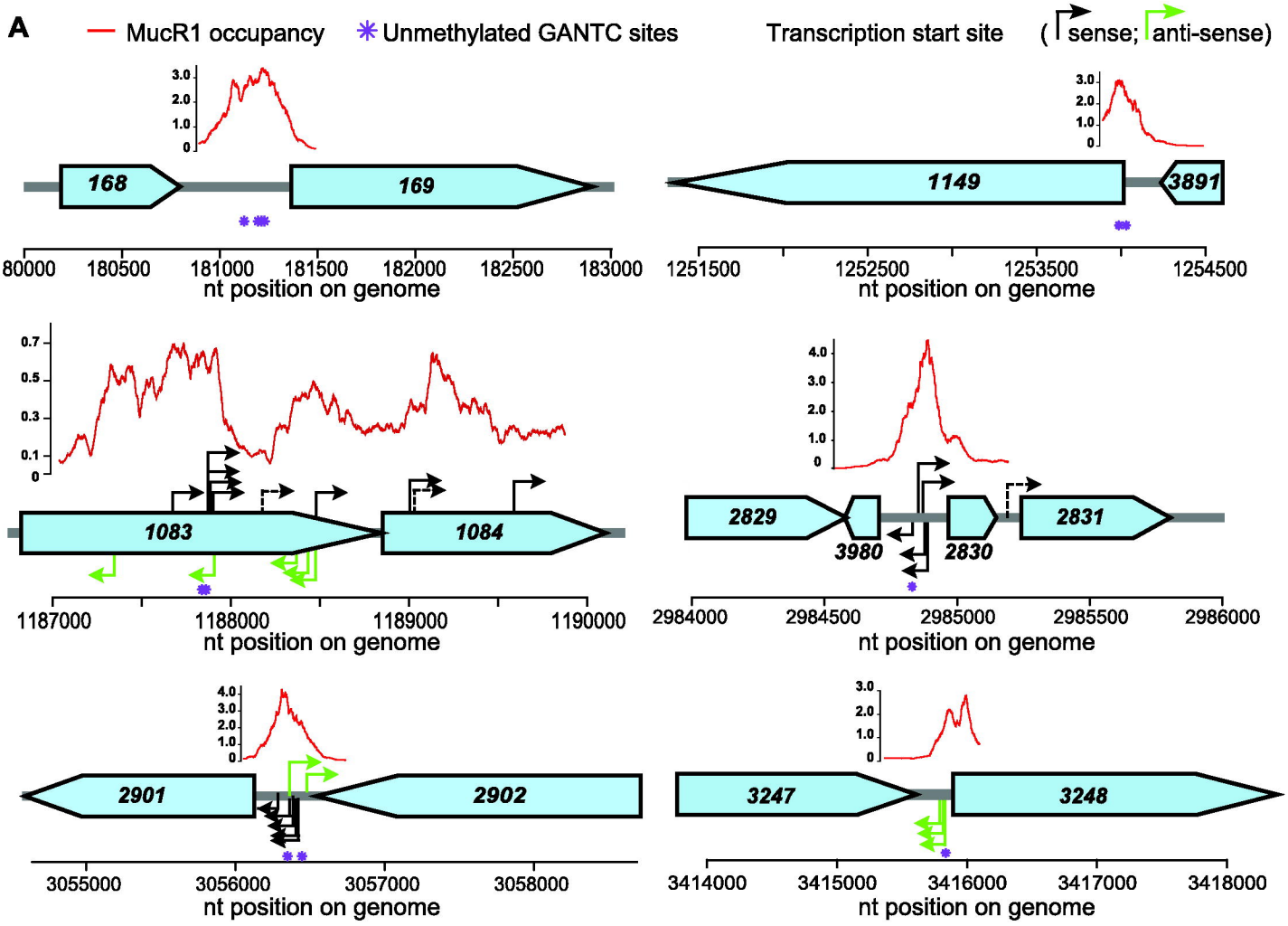
949 *crescentus* and *S. meliloti* genomic DNA.

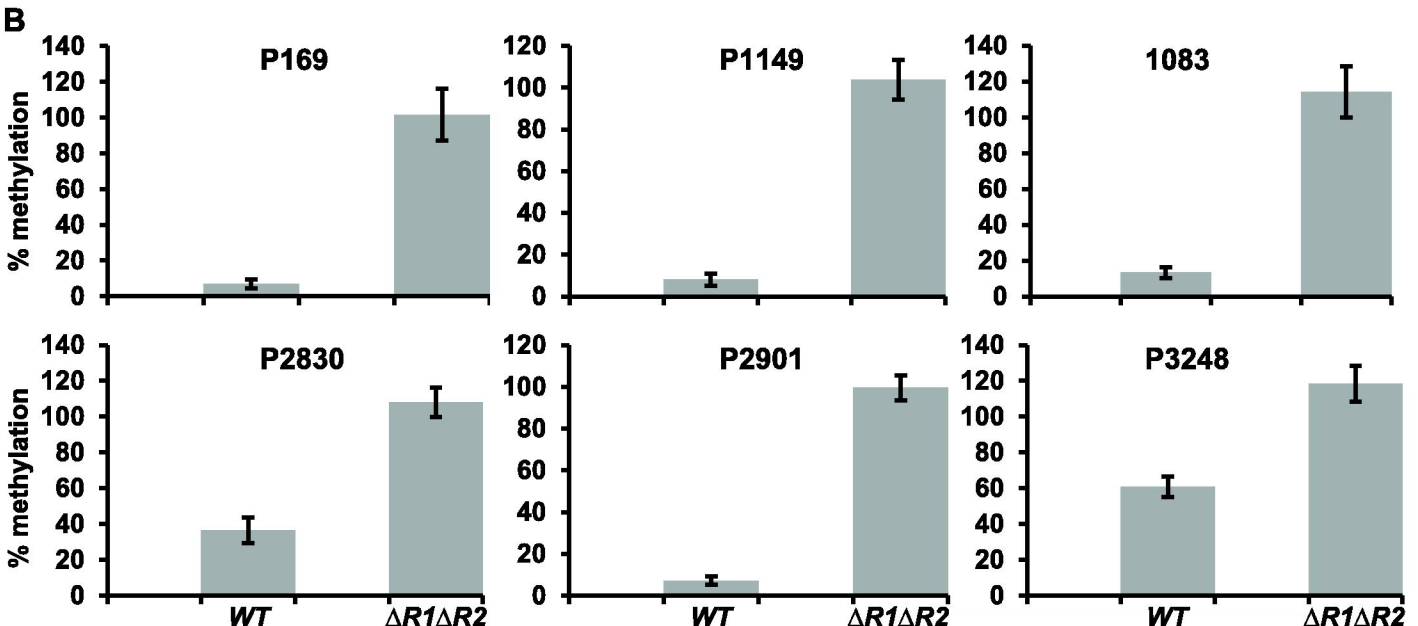
950 **S4 Table.** ChIP-Exo analysis of MucR1 occupancy at different time points (T10, T40, T70, T100)

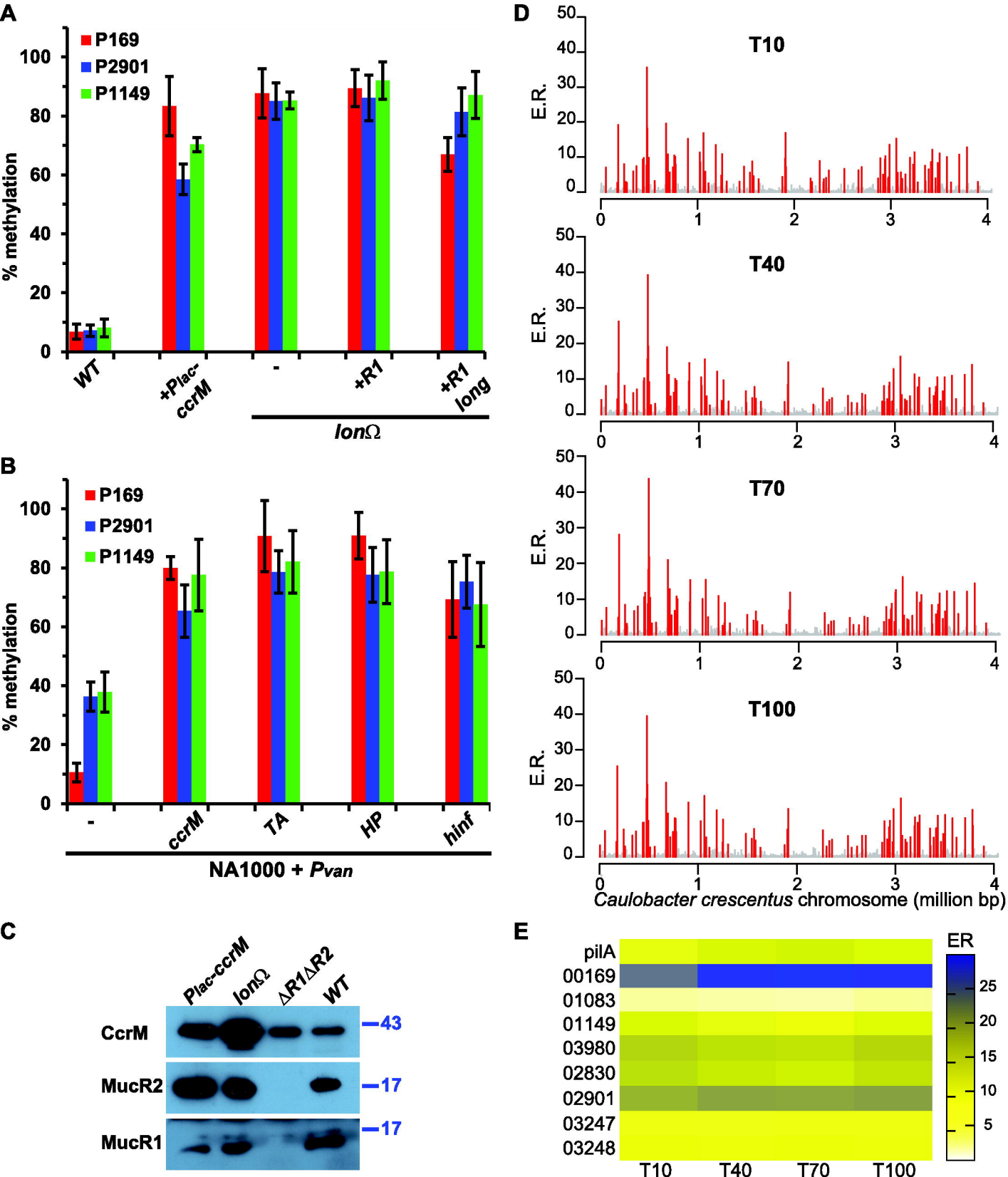
951 during the *C. crescentus* cell cycle.

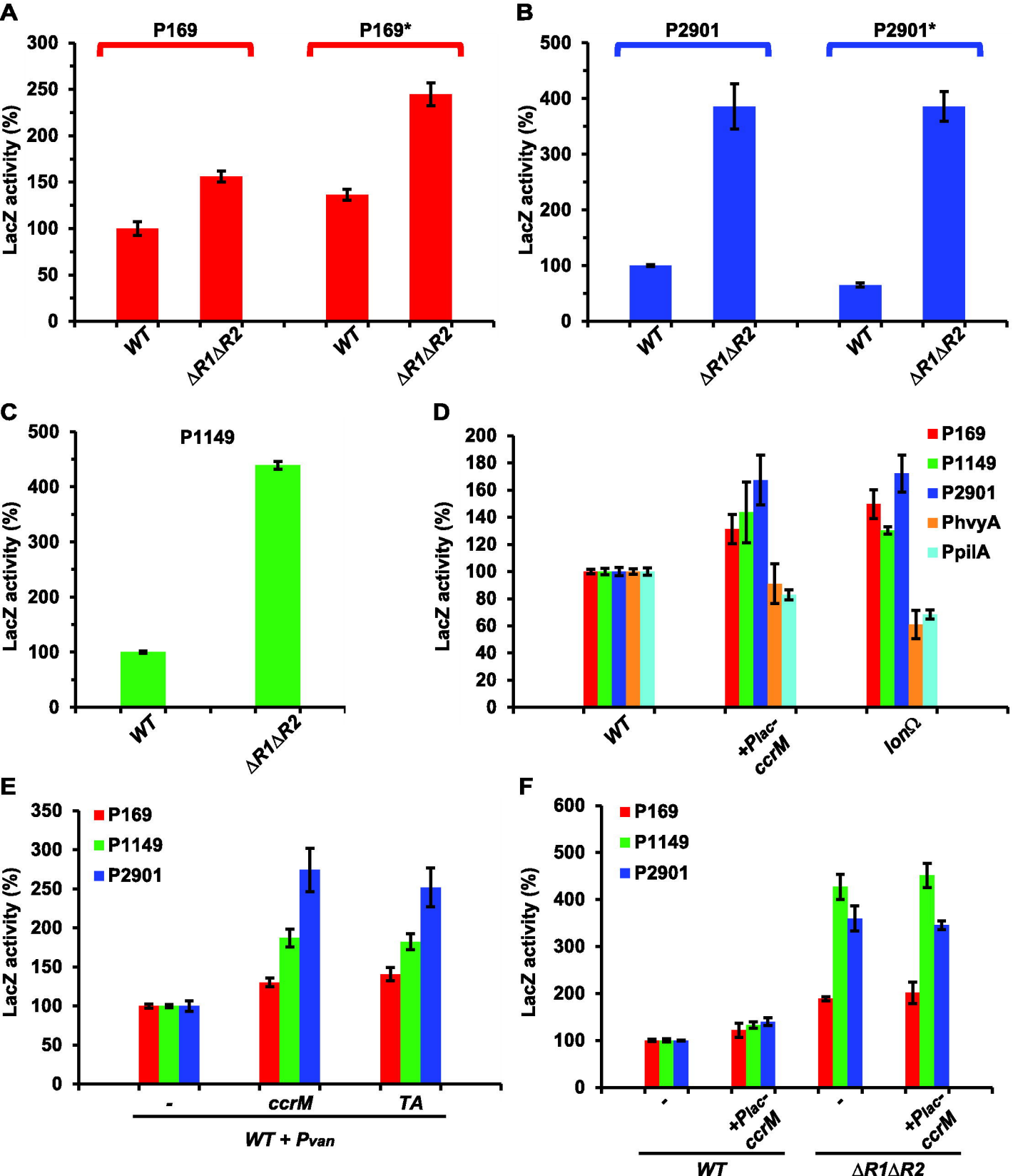
952 **S5 Table.** TSS-EMOTE analysis of transcription start sites in *WT* and Δ*mucR1/2 C. crescentus* cells.

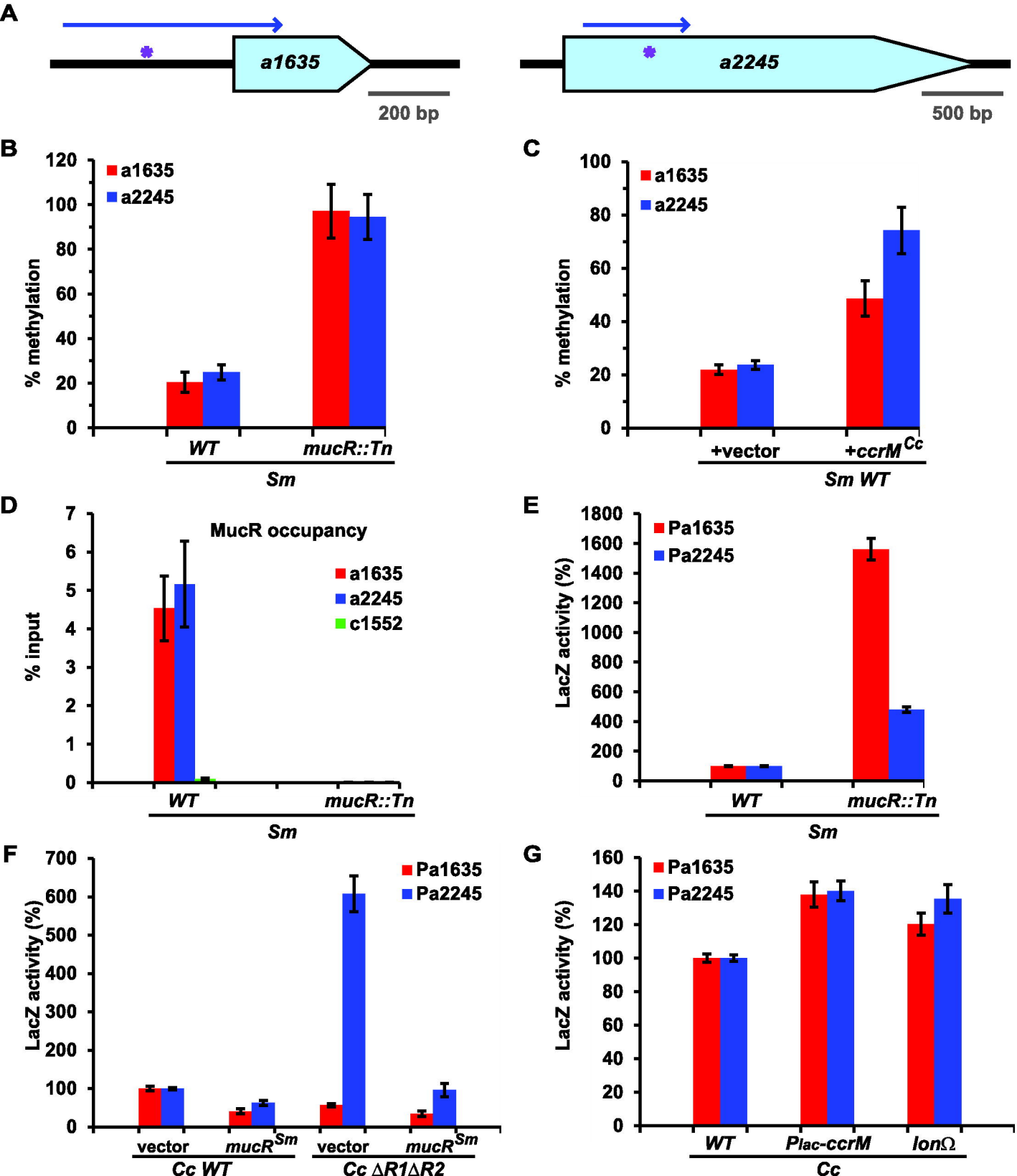

A
MucR1/2
G1 promoters
CtrA~P
Cell cycle
SW (G1 promoters ON)
ST (G1 promoters OFF)
G1
S
G2/Division
B
MucR1/2
m6A
CcrM
Chromosome replication
1
2
3
T10
T40
T70
T100
Origin of replication
Terminus
Replisome
Hemi-methylated chromosome
Fully methylated chromosome

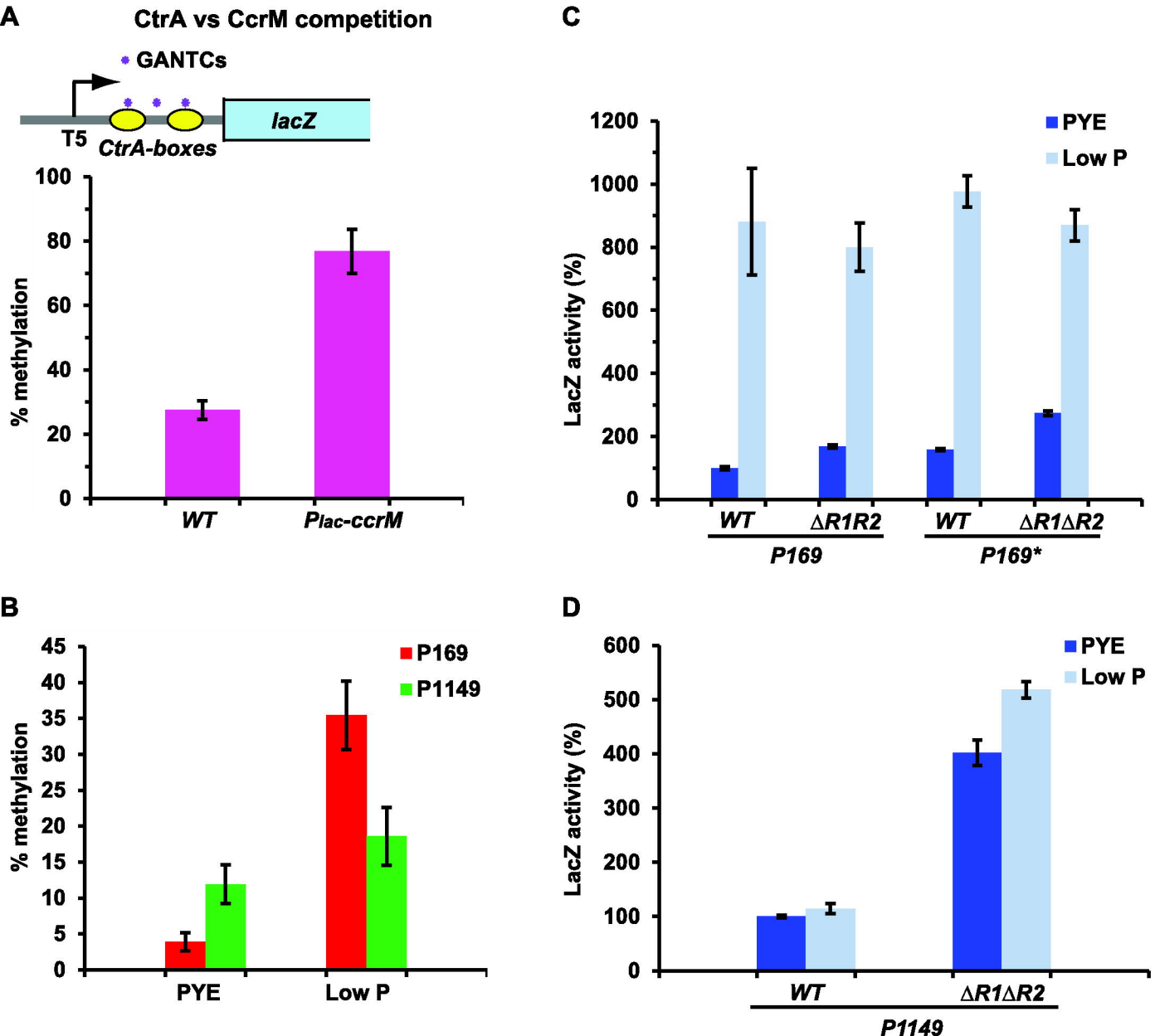

A
MucR1 occupancy
Unmethylated GANTC sites
Transcription start site
(sense; anti-sense)
168
169
180000
180500
181000
181500
182000
182500
183000
nt position on genome
1149
3891
1251500
1252500
1253500
1254500
nt position on genome
1083
1084
1187000
1188000
1189000
1190000
nt position on genome
2829
3980
2830
2831
2984000
2984500
2985000
2985500
2986000
nt position on genome
2901
2902
3055000
3056000
3057000
3058000
nt position on genome
3247
3248
3414000
3415000
3416000
3417000
3418000
nt position on genome


B
% methylation
P169
P1149
1083
P2830
P2901
P3248
WT
ΔR1ΔR2


A
P169
P2901
P1149
% methylation
WT
+Plac-ccrM
-
+R1
+R1 long
lonΩ
B
P169
P2901
P1149
% methylation
-
ccrM
TA
HP
hinf
NA1000 + Pvan
C
Plac-ccrM
lonΩ
ΔR1ΔR2
WT
CcrM
MucR2
MucR1
43
17
17
D
T10
T40
T70
T100
E.R.
Caulobacter crescentus chromosome (million bp)
E
pilA
00169
01083
01149
03980
02830
02901
03247
03248
T10
T40
T70
T100
ER


A
P169
P169*
LacZ activity (%)
300
250
200
150
100
50
0
WT
ΔR1ΔR2
WT
ΔR1ΔR2
B
P2901
P2901*
LacZ activity (%)
500
400
300
200
100
0
WT
ΔR1ΔR2
WT
ΔR1ΔR2
C
P1149
LacZ activity (%)
500
400
300
200
100
0
WT
ΔR1ΔR2
D
P169
P1149
P2901
PhvyA
PpilA
LacZ activity (%)
200
180
160
140
120
100
80
60
40
20
0
WT
+Plac-ccrM
lonΩ
E
P169
P1149
P2901
LacZ activity (%)
350
300
250
200
150
100
50
0
-
ccrM
TA
WT + Pvan
F
P169
P1149
P2901
LacZ activity (%)
600
500
400
300
200
100
0
-
+Plac-ccrM
-
+Plac-ccrM
WT
ΔR1ΔR2


A
a1635
200 bp
a2245
500 bp
B
a1635
a2245
% methylation
WT
mucR::Tn
Sm
C
a1635
a2245
% methylation
+vector
+ccrM Cc
Sm WT
D
MucR occupancy
a1635
a2245
c1552
% input
WT
mucR::Tn
Sm
E
Pa1635
Pa2245
LacZ activity (%)
WT
mucR::Tn
Sm
F
Pa1635
Pa2245
LacZ activity (%)
vector
mucR Sm
vector
mucR Sm
Cc WT
Cc ΔR1ΔR2
G
Pa1635
Pa2245
LacZ activity (%)
WT
Plac-ccrM
lonΩ
Cc


A
CtrA vs CcrM competition
GANTCs
T5
CtrA-boxes
lacZ
% methylation
WT
Plac-ccrM
B
P169
P1149
% methylation
PYE
Low P
C
PYE
Low P
LacZ activity (%)
WT
ΔR1R2
WT
ΔR1ΔR2
P169
P169*
D
PYE
Low P
LacZ activity (%)
WT
ΔR1ΔR2
P1149