Supplementary Methods for

The essential genome of a bacterium

Beat Christen, Eduardo Abeliuk, John M. Collier, Virginia S. Kalogeraki, Ben Passarelli, John A. Coller, Michael J. Fero, Harley H. McAdams, and Lucy Shapiro

(i) *Tn5Pxyl* transposon construction and mutagenesis.

Primers are listed in Supplementary Table 7. The *Tn5* transposase and an adjacent ampicillin resistance cassette was PCR amplified from plasmid pIT2 using primer #1 and #2. Primer #1 incorporated the *Tn5* I-end sequence and a NdeI site. Primer #2 introduced the *Tn5* O-end sequence, the Illumina paired-end adapter sequence (PE1.0, Illumina, Hayward, CA) and a SacI site. The PCR product was cloned into pCR2.1 and further subcloned via NdeI, SacI into pXMCS2 to produce the transposon delivery plasmid pXMCS2*::Tn5Pxyl* (Supplementary Figure 1A). The *Tn5Pxyl* element harbors at one end an outward pointing *Pxyl* promoter. Depending on the orientation of a given *Tn5Pxyl* insertion, the outward facing *Pxyl* promoter will permit expression of adjacent genes.

The transposon delivery plasmid pXMCS2::*Tn5Pxyl* was conjugated from an *E. coli* S17-1 donor strain into a *Caulobacter* Δ*recA* strain. Transposition events were selected onto PYE plates containing kanamycin (20 μg/ml) and nalidixic acid (20 μg/ml) and 0.1% xylose. PYE plates were incubated at 30 ºC for 7 days till kanamycin resistant colonies formed. Pools of approximately 1000-1500 kanamycin resistant colonies were deposited into individual wells of 96-well plates. The resulting *Tn5Pxyl* insertion library contained an estimated 8 x 10^5 transposon mutants.

(ii) DNA library preparation and DNA sequencing

Parallel amplification of DNA fragments covering transposon junctions. DNA fragments that cover transposon junctions from each mutant pool were simultaneously amplified by a semi-arbitrary PCR strategy (Fig. S1). PCR reactions were performed in 96 well format on a DNA Engine Tetrad Thermal Cycler (MJ Research, Waltham, MA) using BioMix® Red (Bioline, Taunton, MA). For each transposon mutant pool 1 μl of a bacterial culture (OD 0.1) was directly added to individual PCR reactions as a template. Terminal adapter sequences compatible to the Illumina sequencing process were amplified from the engineered linker adjacent to the *Tn5Pxyl*

O-end as well as introduced via the PCR primers (#4, #5, #6). In a first-round of PCR, a transposon specific primer #3 that points outwards of the *Tn5Pxyl* element was used in combination with three different semi-arbitrary primers (#4, #5, #6). Each semi-arbitrary primer contained a defined 3' penta-nucleotide sequence, an interspacing 10 bp long arbitrary sequence, and a 5' PE2.0 adapter sequence. The three semi-arbitrary primers possess different 3' pentanucleotide sequences that anneal on average every 300 bp throughout the *Caulobacter* genome. The first round of PCR was performed with the following program: (1) 94˚C for 3 min, (2) 94˚C for 30 s, (3) 42˚C for 30 s, slope -1˚C per cycle, (4) 72˚C for 1 min, (5) go to step 2, 6 times, (6) 94˚C for 30 s, (7) 58˚C for 30 s, (8) 72˚C for 1 min, (9) go to step 6, 25 times, (10) 72˚C for 3 min.

First-round PCR products were further amplified in a second nested PCR step using the Illumina paired-end primer PE1.0 and PE2.0 (#7, #8) and 2 µl of the first round PCR product as template. Primer #7 anneals to the engineered Illumina paired-end sequence adjacent to the transposon O-end. Primer #8 anneals to the 5' tag sequence previously introduced by the first-round semi-arbitrary primers (#4, #5, #6). Second-round PCR was thermo-cycled using the following parameters: (1) 94˚C for 3 min, (2) 94˚C for 30 s, (3) 64˚C for 30 s, (4) 72˚C for 1 min, (5) go to step 2, 30 times, (6) 72˚C for 3 min. Second-round PCR products derived from an estimated 4-5 x 10^5 transposon mutants were pooled according to the semi-arbitrary primer used. DNA fragment of 140 bp-700 bp were size selected and purified by agarose gel electrophoresis prior Illumina sequencing.

Illumina sequencing. Cluster-generation was performed within standard Illumina paired-end flow-cells (Illumina cluster chemistry v2) and settings according to the Illumina cluster generation protocol. The DNA sample concentration of each transposon junction library was titrated to produce approximately 1.3 x 10^5 clusters per tile on the flow cell. Paired-end sequencing with 76 bp read length from both ends of a given DNA fragment was performed on an Illumina GAIIx instrument. Sequencing chemistry v4 and standard paired-end sequencing primers (PE 1.0 and PE 2.0, Illumina) were used.

(iii) Sequence processing

Sequence analysis. Base-calling from raw images was performed using the genome analyzer software suite OLB version 1.6 (Illumina). A *phiX* control lane was specified to calculate

crosstalk matrixes and calibrate phasing parameters. Sequencing strings reading out of the *Tn5Pxyl* transposon elements were identified by requiring a perfect match to the last 15 bases of the O-end sequence (GTGTATAAGAGTCAG). Paired-end reads were further processed to remove the 19 bp long *Tn5Pxyl* O-end sequence and the 10 bp long arbitrary sequence of the semi-arbitrary primers. The resulting reads contained *Caulobacter* specific DNA strings that were subsequently mapped onto the *Caulobacter crescentus* NA1000 reference chromosome (Marks et al, 2010) using the Maq software (Li et al, 2008). A SNPs and Indels analysis was performed to identify any sequencing error sites within the reference *Caulobacter crescentus* NA1000 genome. Sequencing error could lead to some small sites being incorrectly deemed essential. The SNP and Indel analysis showed that the genome sequence of the *Caulobacter* Δ*recA* strain used in this study was identical to the published reference genome sequence *of Caulobacter crescentus* NA1000 with the exception of the engineered 700bp deletion covering part of the *recA* gene. The following read alignment criteria were used to identify genome-wide transposon insertion sites: (a) A perfect match onto the reference genome was required for the first 25 bp of each read and its mate. (b) Each read and its mate had to be located in correct paired-end orientation less than 700 bp apart. Upon insertion into the target DNA, *Tn5* tranposition generates a nine base pair long duplication, which was taken into account for subsequent insertion site analysis. We defined the location of a given transposon insertion as the genome position of the first reference base detected immediately after reading out of the transposon I-end. The I-end contains the outward pointing *Pxyl* promoter. For every transposon insertion located within a coding region, we determined the insertion frame. The *Pxyl* promoter of the transposon element only permits transcription of downstream sequences for insertions located in the sense direction (frames +1, +2, +3). In addition, the *Tn5Pxyl* element carries an engineered ribosome binding site and an internal ATG codon immediately adjacent to the *Pxyl* promoter element. Insertions located in the +1 frame initiate translation at the internal ATG start codon and read through the I-end sequence into the downstream coding sequence.

(iv) Essentiality analysis.

Identification of non-disruptable non-coding genome elements. The frequency of transposon insertion within non-coding sequences of the *Caulobacter* genome was analyzed and non-disruptable DNA regions at least 90 bp long were identified. Non-disruptable non-coding DNA regions located within the essential promoter region of the essential genes were excluded from

further analysis. The boundaries of the non-disruptable DNA elements detected were corrected to compensate for the 9 bp duplication generated upon *Tn5* transposition.

High-resolution mapping of all essential protein coding sequences. We determined the number of transposon insertions disrupting the coding region for every ORF. Transposon insertions in the stop codon of an ORF or located in the +1 reading frame were excluded. Insertions targeting the stop codon of an ORF will not abrogate protein production while insertions located in the +1 frame will permit translation of downstream sequences due to the engineered RBS within the transposon element. Transposon insertions in the +1 frame of a target ORF do not necessarily abrogate protein function and could yield functional fusion proteins or split a given protein into two polypeptides fragments that retain the capability to assemble into a function protein. Sense insertions in the last 9 bp of an ORF and anti-sense insertions targeting the first 9 bp were omitted since these insertions can still provide a full-length copy of the ORF due to the 9 bp duplication generated upon *Tn5* transposition. The average density of transposon insertions was calculated for every annotated ORF by dividing the number of disruptive transposon insertions by the target length of the ORF. The target length was defined as the length of the ORF in base pairs excluding the stop codon and subtracting 9 bp due to the duplication generated by *Tn5* transposition. In addition, we determined for every ORF the length of the non-disrupted 5' region and calculated the distance between the start codon and the first not in-frame insertion event downstream. The measurement was considered robust if a second insertion event was detected no further than 100 bp apart from of the first internal insertion event. Otherwise the first insertion event was considered as noise and the length of the non-disrupted 5' region was calculated as the distance between the start codon and the second insertion event. Furthermore, for every coding region we calculated the length of the largest internal non-disruptable coding segment detected. Similar robustness criteria as for the determination of the non-disrupted 5' region were applied. *Caulobacter* ORFs were classified as essential if the non-disruptable 5' region covered more than 60% of the total ORF length. In order to identify essential ORFs that might be mis-annotated we also classified ORFs as essential if the ORF showed an insertion density of less than one insertion in 50 base pairs and the biggest non-disruptable internal gap identified covered more than 60% of the total ORF length. The start codons of essential ORFs were re-annotated if anti-sense insertions were detectable within 5' regions. We chose the closest ATG, GTG or TTG codon positioned downstream of an anti-sense insertions as the correct start

codon of mis-annotated essential ORFs. *Caulobacter* ORFs that showed an insertion density of less than one insertion every 25 bp and harbored non-disruptable 5' region covering less than 60% of the total ORF length were classified as non-essential ORFs with high fitness costs.

Essential promoter and operon analysis. We found that the orientation of transposon insertions was significantly biased within the upstream regions of essential genes. Transposon insertions within the promoter region of an essential gene are viable if the transposon specific *Pxyl* promoter points in sense direction (sense insertion) and rescues transcription. Conversely, transposon insertions in the opposite orientation (anti-sense insertions) are lethal because they abrogate transcription of the essential protein or RNA coding sequences downstream. For each essential ORF we identified the corresponding essential promoter region by measuring the distance between the annotated start codon of the ORF and the first anti-sense insertion upstream. Similarly, we mapped the essential promoter region of essential RNAs by measuring the distance between the annotated start of the RNA element and the first anti-sense insertion detected upstream. ORFs harboring essential promoter regions that fully included one or more upstream ORFs were classified as essential operons.

LacZ-reporter assays. The promoter region and translational start site of *chpT* was determined using a series of translational *lacZ*-reporter fusions. DNA regions extending 156 bp, 89 bp, 87 bp, 73 bp or 70 bp upstream of the re-annotated start codon of *chpT* (Met_{29} codon in the old mis-annotated *chpT* coding sequence) and covering the first 20 amino acids of the re-annotated coding sequence of *chpT* were cloned into the BglII and SpeI site of the *lacZ*-reporter plasmid pR9TT. The reporter constructs were transformed into *Caulobacter* wild type cells. The LacZ-activity of all reporter constructs was determined in triplicates using standard ONPG based β-galactosidase assays and compared to the basal LacZ-activity of an empty pR9TT control plasmid.

DNA sequences covering each essential promoter region listed in Supplementary Table 6 and extending 40 amino acids into the coding region of the corresponding gene were PCR amplified and cloned via BglII and SpeI sites into the *lacZ* reporter plasmid pR9TT. The LacZ-activity of the resulting translational *lacZ*-reporter constructs were assayed in *Caulobacter* wild type cells using standard ONPG based β-galactosidase assays. The β-galactosidase activities reported in

Supplementary Table 6 represent the average of at least three independent measurements derived from mid-log phase *Caulobacter* cultures.

Targeted deletion analysis. For a subset of identified essential genes encoding hypothetical proteins we attempted targeted deletions via a two-step homologous recombination strategy (Supplementary Table 3). DNA cassettes covering either 700 bp of the upstream or downstream DNA regions of the essential ORF of interest were PCR amplified from genomic *Caulobacter* DNA and spliced together by a subsequent SOE-PCR step. The resulting PCR products were cloned into the EcoRI and SpeI site of pNPTS138. The pNPTS138 derivatives were conjugated into *Caulobacter* wild type strains and site-specific integrants were selected on PYE plates containing kanamycin (20 μg/ml). The pNPTS138 plasmid carries the *sacB* gene that confers sucrose sensitivity. Second recombination events were selected on PYE plates supplemented with 0.3% sucrose. If the gene under investigation is essential, the sucrose counter-selection will only yield wild type revertants or *sacB* null mutations while in-frame deletion events will not be recovered. Second recombinants were arrayed in 96-well plates and genotyped by PCR on a DNA Engine Tetrad Thermal Cycler (MJ Research, Waltham, MA) using BioMix® Red (Bioline, Taunton, MA).

Hierarchical clustering and cell-cycle dependent expression analysis. For every essential ORF, we examined the cell-cycle dependent expression profile of the corresponding mRNA transcript as previously reported (McGrath et al, 2007). We computed a "cell cycle regulation index" (de Lichtenberg et al, 2005) for each one of these mRNA profiles. A total of 84 essential ORFs were classified as cell-cycle regulated. For further clustering analysis, we focused on expression profiles of cell-cycle regulated mono-cistronic genes or lead genes of operon. Expression profiles of these essential ORFs were subsequently clustered via a bottom-up hierarchical clustering approach (Bioconductor R package) using the Pearson correlation distance and complete linkage for measuring inter-cluster distances.

Phylogenetic conservation analysis of essential ORFs. For each essential protein we calculated a bit-homology score based on the sequence conservation according to the STRING database (von Mering et al, 2003). Bit-homology scores of the *Caulobacter* essential ORFs were calculated to determine the phylogenetic conservation across the proteobacteria, and then hierarchical clustering was used to group the essential ORFs according to their phylogenetic

distribution. The essential gene sets for *E. coli* was defined using experimental gene essentiality data sets derived from PEC (profiling of *E. coli* chromosome) (Gerdes et al, 2003; Yamazaki et al, 2008). If both data sets disagreed on essentiality the Keio data set (Baba et al, 2006) was used instead for classification.

(v) Statistical data analysis.

Test for uniform distribution of transposon insertion sites. Only insertion sites within nonessential ORFs were considered in this analysis. To dismiss boundary effects due to nearby essential genome features, we excluded transposon insertion sites within the first or last 10 percent of nonessential ORFs. If insertion is equal probable at all sites, then the distances between consecutive insertion sites (insertion gaps) should be exponentially distributed. Using a Kolmogorov-Smirnov test we found that the chromosomal distribution of insertion gaps had an exponential distribution with insertion probability of 0.1 per base pair (accepting the null hypothesis with p-value of 0.4).

Calculation of p-values for essentiality. Next, assuming uniform *Tn5Pxyl* hit rates across the entire chromosome and neutral fitness costs for every possible insertion generated at any genome location, we calculated p-values for each non-disruptable genome region. P-values were calculated according to

$$p_{(l)} = \left(1 - \frac{l}{g_{size}}\right)^{n_{Tn5}}$$

where, $p_{(l)}$ is the p-value to observe no *Tn5* insertion events within a genome region of length l or bigger; g_{size} is the size of the target genome in base pairs (4,042,929 bp); l is the length of the non-disruptable genome region in base pairs; n_{Tn5} is the estimated number of independent *Tn5* insertions present within the entire transposon mutant library.

n_{Tn5} was approximated according to:

$$n_{Tn5} = \frac{\log_{10}\left(1 - \frac{n_{mapped}}{2g_{size}}\right)}{\log_{10}\left(1 - \frac{1}{2g_{size}}\right)}$$

where n_{mapped} is the number of unique *Tn5* insertion sites mapped (428,735).

Transposon orientation analysis. The statistical significance of the observed transposon orientation bias within upstream regions of essential genes was assessed by a Mann-Whitney U test. For each gene we defined the CSI value as the distance between the start codon and the closest sense insertion upstream of the gene. Similarly, we defined the CAI value as the distance between the start codon and the closest anti-sense insertion upstream of the gene. For the group of essential genes, we compared the two distributions of CSI and CAI values and found that these two distributions differed significantly (median CSI -44 bp, median CAI -14 bp, Mann–Whitney two-tailed, U = 20436.5, n1 = n2 = 327, p-value=2.6×10^{-17}). As a control, we repeated the same analysis but considered only non-essential genes.

The size distribution of essential promoter regions belonging to cell-cycle regulated or non-cell-cycle regulated genes were statistically analyzed by a Mann-Whitney test and differed significantly (U = 1595.5, n1 = 45, n2 = 242, p-value 0.0018).

Enrichment of essential ORFs along the chromosome. Across the entire *Caulobacter* and *E. coli* genomes, we counted the total number of ORFs (n) and the number of essential ORFs (k) found within a 500 kbp moving window (window step size 10kbp). We also determined m, the total number of essential ORFs, and N, the total number of ORFs encoded within the entire *Caulobacter* or *E.coli* genome. Using a hypergeometric distribution, we calculated the probability E(n,k,m,N) of observing k or more essential ORFs within a given genomic window and assigned p-values. P-values were calculated across the whole genome using a window step size of 10kbp and plotted as –log(p-value) (Fig. 3A).

Supplementary Figures and tables:

Sup. Fig. 1. Genomic high-resolution transposon scanning. (**A**). A feature map of the engineered pXMCS2::*Tn5Pxyl* transposon delivery plasmid is shown. (**B**). The nested semi-arbitrary PCR strategy used for simultaneous amplification of transposon junctions is shown. Terminal adapters (blue and orange) compatible for Illumina paired-end sequencing are introduced by the arbitrary PCR primers or via linker-sequence adjacent to the transposon O-end. The second-round PCR products are sequenced from both ends using standard paired-end sequencing primers. (**C**). Base-call frequencies are shown for a representative paired-end sequencing experiment. The 19 bp long transposon O-end as well as the defined penta-nucleotide of the arbitrary primer produce discrete base calls.

Sup. Fig. 2. Experimental and theoretical distribution of transposon insertions along the *Caulobacter* genome. (**A**) The fraction of the *Caulobacter* genome for which the observed spacing between consecutive *Tn5* elements does not exceed a given distance in bp is shown. (**B**). The distance between consecutive transposon insertions is plotted against their frequency for the experimental data set (red dots). For small gap sizes (<60-70 bp) the experimental data set follows the theoretical probability function assuming a uniformly distributed insertion frequency across the entire *Caulobacter* genome. Large gap sizes observed within the experimental data set indicate non-disruptable genome elements.

Sup. Fig. 3. Transposon insertion density across non-essential and essential genes. The insertion frequency across normalized gene length is uniform for non-essential genes (blue bars) while essential genes (red bars) tolerate insertions mainly toward the 3' end of the gene.

Supplemental Table 1. Essential *Caulobacter* ORFs with mis-annotated translational start sites

Gene_ID	Start[a]	Stop[b]	Diff. [AA][c]	Functional Annotation
CCNA_00052	55611	55231	31	Predicted metal-dependent hydrolase
CCNA_00100	111638	110979	7	Ribulose-phosphate 3-epimerase
CCNA_00211	227686	228693	35	Agmatine deiminase
CCNA_00217	234091	236127	38	Thiol:disulfide interchange protein DsbD
CCNA_00268	282130	283956	15	DNA polymerase III subunit gamma/tau
CCNA_00469	483454	482231	52	Glycosyltransferase
CCNA_00748	805230	805484	23	Ferrous iron transport protein A
CCNA_00976	1056336	1055002	150	Signal peptide peptidase
CCNA_01304	1431129	1431329	44	Hypothetical protein
CCNA_01676	1799591	1803712	25	Hypothetical protein
CCNA_01952	2096903	2095782	22	N-acetylmuramoyl-L-alanine amidase
CCNA_02020	2167827	2167336	30	NADH-quinone oxidoreductase chain I
CCNA_02056	2205949	2205383	53	Chaperone, ATP12 family protein
CCNA_02299	2445860	2446528	13	Cell division ATP-binding protein ftsE
CCNA_02320	2463195	2462335	79	Deoxycytidine triphosphate deaminase
CCNA_02964	3119953	3120858	47	Thioredoxin reductase
CCNA_03378	3550841	3551641	39	Myo-inositol-1(or 4)-monophosphatase
CCNA_03527	3684726	3683461	32	Glutamate-cysteine ligase
CCNA_03534	3692947	3693894	13	4-hydroxybenzoate polyprenyltransferase
CCNA_03584	3739999	3740676	28	Histidine phosphotransferase chpT
CCNA_03664	3825012	3825773	15	Dihydrodipicolinate reductase
CCNA_03731	3899293	3898748	21	Histidinol-phosphatase
CCNA_03732	3899494	3900552	38	Mannose-6-phosphate isomerase
CCNA_03735	3904425	3906137	101	Transketolase
CCNA_03755	3925947	3924496	20	D-glycero-D-manno-heptose-7-phosphate 1-kinase
CCNA_03827	3992475	3993671	14	Phosphopantothenate-cysteine ligase
CCNA_03837	4003055	4002543	11	Hypothetical protein
CCNA_03842	4009598	4008861	11	Molybdopterin biosynthesis MoeB protein
CCNA_03859	4022883	4022266	27	Two-component response regulator cenR
CCNA_03879	4042684	4041629	12	Uroporphyrinogen decarboxylase

[a] Genome coordinates of the re-annotated translational start sites are shown.
[b] Genome coordinates for the end position of ORFs are shown.
[c] Difference between the old and new ORF annotation in amino acids.

Supplemental Table 2. Essential *Caulobacter* ORFs of unknown function

Gene_ID	Functional Annotation	Length [AA]	Protein Families & Domains[a]
CCNA_00052	predicted metal-dependent	126	UPF0054 superfamily
CCNA_00063	virulence factor MviN	520	
CCNA_00107	hypothetical protein	66	Trm112p superfamily
CCNA_00306	hypothetical protein	238	DUF374
CCNA_00308	hypothetical protein	222	PTPc superfamily
CCNA_00309	cytosolic protein	78	
CCNA_00354	hypothetical protein	236	NLPC_P60 superfamily
CCNA_00380	cytosolic protein	180	DUF721
CCNA_00524	cytosolic protein	375	DUF185
CCNA_00546	hypothetical protein	261	
CCNA_00735	hypothetical protein	190	DUF3035
CCNA_00761	hypothetical protein	70	
CCNA_01118	hypothetical protein	110	DUF1491
CCNA_01211	hypothetical protein	292	
CCNA_01304	hypothetical protein	67	
CCNA_01607	hypothetical protein	240	SNARE_assoc superfamily
CCNA_01614	hypothetical protein	172	
CCNA_01676	hypothetical protein	1399	DUF490
CCNA_01724	hypothetical protein	251	DUF2133
CCNA_01725	PQQ repeat protein	477	
CCNA_01726	GTP-binding protein	588	Ras_like_GTPase superfamily
CCNA_01761	hypothetical protein	369	YjgP_YjgQ superfamily
CCNA_01762	hypothetical protein	380	YjgP_YjgQ superfamily
CCNA_01921	TPR repeat protein	98	Sdh5 superfamily
CCNA_01935	cytosolic protein	118	DUF59 superfamily
CCNA_01968	hypothetical protein	290	Pentapeptide superfamily
CCNA_01987	hypothetical protein	281	DUF1009
CCNA_01991	hypothetical protein	213	OmpH superfamily
CCNA_02063	lipoprotein, ComL family	309	
CCNA_02094	NAD-dependent oxidoreductase	305	NADB_Rossmann superfamily
CCNA_02528	hypothetical protein	219	DUF205
CCNA_02841	hypothetical protein	80	
CCNA_02942	hypothetical protein	390	DUF1022
CCNA_03213	putative polyhydroxyalkanoic	102	PHA gran_rgn superfamily
CCNA_03274	hypothetical protein	39	
CCNA_03307	hypothetical protein	89	
CCNA_03401	cytosolic protein	238	DUF1013
CCNA_03405	hypothetical protein	226	COQ9
CCNA_03428	cytosolic protein	90	DUF2312
CCNA_03630	hypothetical protein	189	GepA superfamily
CCNA_03640	ferredoxin NAD(+) reductase	425	Pyr_redox superfamily
CCNA_03648	ATP/GTP hydrolase	148	UPF0079 superfamily
CCNA_03714	ABC transporter ATP-binding	252	P-loop NTPase superfamily
CCNA_03715	yhbN precursor	187	OstA superfamily
CCNA_03716	hypothetical protein	221	DUF1239
CCNA_03780	hypothetical protein	248	DUF1223
CCNA_03837	hypothetical protein	182	SH3_3 superfamily
CCNA_03866	hypothetical protein	167	DUF2159
CCNA_03877	hypothetical protein	172	UPF0093

[c] Conserved protein domains as listed in the Conserved Domain Database (Marchler-Bauer et al, 2011) are shown.

Supplemental Table 3. Attempted deletion for a subset of essential *Caulobacter* ORFs of unknown function.

Gene_ID	Functional Annotation	2nd Recombinant[a]		
		wt	Δ	sacB $^-$
CCNA_00107	hypothetical protein	37	0	10
CCNA_00306	hypothetical protein	21	0	27
CCNA_00354	hypothetical protein	40	0	6
CCNA_00380	cytosolic protein	88	0	6
CCNA_01118	hypothetical protein	35	0	13
CCNA_01607	hypothetical protein	38	0	8
CCNA_01724	hypothetical protein	23	0	25
CCNA_03405	hypothetical protein	24	0	24
CCNA_03428	hypothetical protein	41	0	7
CCNA_03630	hypothetical protein	37	0	11
CCNA_03780	hypothetical protein	21	0	27

[a] Second recombination genotypes observed after the sucrose counter selection step are shown. Only revertants to wild type (wt) or *sacB* mutations (SacB$^-$) were detectable while in-frame deletions of the probed hypothetical proteins were not recovered.

Supplemental Table 4. Essential *Caulobacter* cell-cycle regulators, transcription factors and sigma factors

Gene_ID	Functional Annotation	Reference
CCNA_03130	ctrA, cell cycle response regulator	(Quon et al, 1996)
CCNA_02328	gcrA, cell cycle regulatory protein	(Holtzendorff et al, 2004)
CCNA_00008	dnaA, chromosomal replication initiator protein	(Gorbatyuk & Marczynski, 2001)
CCNA_00382	ccrM, adenine-specific methyltransferase	(Stephens et al, 1996),
CCNA_00948	sciP, transcriptional regulator	(Gora et al, 2010; Tan et al, 2010)
CCNA_00684	chrR, transcriptional activator	(Lourenco & Gomes, 2009)
CCNA_03142	rpoD, RNA polymerase sigma factor	(Malakooti & Ely, 1995)
CCNA_03195	rpoH, RNA polymerase sigma-32 factor	(Wu & Newton, 1997)
CCNA_00690	CarD-like transcriptional regulator	This work
CCNA_01864	transcriptional regulator, TetR family	This work
CCNA_02014	biotin operon repressor	This work
CCNA_03726	transcriptional regulator	This work
CCNA_03862	regulatory protein tenI	This work

Supplemental Table 5. TF binding sites and predicted motifs within essential promoter regions:

Regulator	Motif[a]	Essential promoter[b]
CtrA, cell cycle response regulator	104, 136	RR_03130 (*ctrA*)
	59	RR_00382 (*ccrM*)
	41	RR_00948 (*sciP*)
	51	RR_01952 (*amiC*)
	27	RR_02246 (*mipZ*)
	35	RR_02547 (*divK*)
	17	RR_02623 (*ftsZ*)
	38	RR_03819 (*ftsK*)
DnaA, chromosomal replication initiator protein	112	RR_01535 (*ssb*)
	80	RR_01737 (*dnaB*)
	101	RR_02328 (*gcrA*)
	20	RR_00948(sciP)
SciP, transcriptional regulator	68	RR_00382 (*ccrM*)
	170	RR_03130 (*ctrA*)
	35, 94, 104	RR_03754 (*fzlA*)
CcrM, adenine-specific methyltransferase	21	RR_00261 (*nrfF*)
	85, 281	RR_00530 (*rplJ*)
	39, 90	RR_01304
	43	RR_01535 (*ssb*)
	10	RR_01590 (*ligA*)
	118	RR_02328 (*gcrA*)
	9	RR_02543 (*rpmG*)
RpoD, RNA polymerase sigma factor[c]	34	RR_01770 (*ndk*)
	41	RR_00691 (*fdyA*)
	41	RR_02194 (*fumC*)
	42	RR_03006 (*nadA*)
	43	RR_03582 (*cysQ*)
	50	RR_03729
Unknown regulator for motif cc_01[c]	67	RR_03195 (*rpoH*)
	57	RR_01968
Unknown regulator for motif cc_04[c]	74	RR_03195 (*rpoH*)
	67	RR_01968

[a] The position of predicted and biochemically identified motifs within the essential promoter region of cell-cycle regulated genes are indicated. Motif coordinates are shown in bp upstream of translational start sites in Supplemental Data, DT2.

[b] Motif contained within essential promoter region of cell-cycle regulated essential genes are listed. The CcrM "motifs" are the CcrM methylation sites.

[c] Predicted motif as identified by (McGrath et al, 2007).

Supplemental Table 6. Promoter activity of various essential promoter regions:

Gene_ID	Functional Annotation	Essential promoter[a]	Activity[b] (Miller Units)
CCNA_01395	Electron transfer flavoprotein	14 bp	20 ± 1
CCNA_03628	Histidyl-tRNA synthetase	14 bp	148 ± 7
CCNA_03719	Undecaprenyl-diphosphatase	15 bp	24 ± 1
CCNA_01998	Uridylate kinase	18 bp	29 ± 1
CCNA_00520	Ribose-phosphate pyrophosphokinase	19 bp	11 ± 2
CCNA_00758	Protein translation Elongation factor P EF-P	20 bp	110 ± 2
CCNA_01950	Bacterial peptide chain Release factor 2 (RF-2)	20 bp	4 ± 1
CCNA_03341	TolQ protein	21 bp	10 ± 1
CCNA_00116	Phosphoglucosamine mutase	24 bp	47 ± 2
CCNA_03781	Aconitate hydratase	27 bp	77 ± 2
CCNA_00492	Cysteinyl-tRNA synthetase	28 bp	28 ± 1
CCNA_01547	Methionyl-tRNA synthetase	30 bp	63 ± 11
CCNA_03717	Ribonuclease D	30 bp	23 ± 2
CCNA_00062	Tryptophanyl-tRNA synthetase	31 bp	100 ± 1
CCNA_03703	Cytidylate kinase	43 bp	4 ± 1
CCNA_01770	Nucleoside diphosphate kinase	44 bp	125 ± 4
CCNA_00261	Ribonucleoside-diphosphate reductase β-subunit	46 bp	127 ± 4
CCNA_01253	Dihydrodipicolinate synthase	50 bp	1251 ± 10
CCNA_00346	Dihydrolipoamide dehydrogenase	53 bp	363 ± 16
CCNA_02008	Prolyl-tRNA synthetase	64 bp	181 ± 4
CCNA_02064	N-acetylglucosamine deacetylase	65 bp	545 ± 10
CCNA_02063	Lipoprotein, ComL family	67 bp	332 ± 10
CCNA_03384	LSU ribosomal protein L31P	69 bp	18 ± 1
CCNA_03428	Cytosolic protein	69 bp	1256 ± 54
CCNA_00899	Bacterial peptide chain Release factor 1 (RF-1)	87 bp	15 ± 1
CCNA_00737	Isoleucyl-tRNA synthetase	124 bp	9 ± 1
control	empty pR9TT vector	NA	0 ± 0.7

[a] Length of mapped essential promoter region

[b] Promoter activity (Miller units) of translational LacZ reporter fusions were determined in triplicates. ± indicates the standard error of the mean for each measurement.

Supplemental Table 7. Primer list:

Primer	Oligonucleotide Sequence
#1	CATATGGGAGATCTGATCAAGAGACAGGTCGACCGATCCCGTACACAAGTAGC
#2	AGAGCTCGCGTCACACTCATCGGTTGGGTGACACTCTTTCCCTACACGACGCTCTTCCGATC-TACTTGTGTATAAGAGTCAGTTACC
#3	GCCAGGGTTTTCCCAGTCACGA
#4	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNACGCC
#5	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNCCTGG
#6	CTCGGCATTCCTGCTGAACCGCTCTTCCGATCTNNNNNNNNNNCCTCG
#7	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT
#8	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT

References

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* **2:** 2006 0008

de Lichtenberg U, Wernersson R, Jensen TS, Nielsen HB, Fausboll A, Schmidt P, Hansen FB, Knudsen S, Brunak S (2005) New weakly expressed cell cycle-regulated genes in yeast. *Yeast* **22:** 1191-1201

Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, Bhattacharya A, Kapatral V, D'Souza M, Baev MV, Grechkin Y, Mseeh F, Fonstein MY, Overbeek R, Barabasi AL, Oltvai ZN et al (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J Bacteriol* **185:** 5673-5684

Gora KG, Tsokos CG, Chen YE, Srinavasan BS, Perchuk BS, Laub MT (2010) A cell-type-specific protein-protein interaction modulates transcriptional activity of a master regulator in *Caulobacter crescentus*. *Mol Cell*

Gorbatyuk B, Marczynski GT (2001) Physiological consequences of blocked *Caulobacter crescentus dnaA* expression, an essential DNA replication gene. *Mol Microbiol* **40:** 485-497

Holtzendorff J, Hung D, Brende P, Reisenauer A, Viollier PH, McAdams HH, Shapiro L (2004) Oscillating global regulators control the genetic circuit driving a bacterial cell cycle. *Science* **304:** 983-987

Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research* **18:** 1851-1858

Lourenco RF, Gomes SL (2009) The transcriptional response to cadmium, organic hydroperoxide, singlet oxygen and UV-A mediated by the sigmaE-ChrR system in *Caulobacter crescentus*. *Mol Microbiol* **72:** 1159-1170

Malakooti J, Ely B (1995) Principal sigma subunit of the *Caulobacter crescentus* RNA polymerase. *J Bacteriol* **177:** 6854-6860

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL et al (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic acids research* **39:** D225-229

Marks ME, Castro-Rojas CM, Teiling C, Du L, Kapatral V, Walunas TL, Crosson S (2010) The genetic basis of laboratory adaptation in *Caulobacter crescentus*. *Journal of bacteriology* **192:** 3678-3688

McGrath PT, Lee H, Zhang L, Iniesta AA, Hottes AK, Tan MH, Hillson NJ, Hu P, Shapiro L, McAdams HH (2007) High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons. *Nat Biotechnol* **25:** 584-592

Quon KC, Marczynski GT, Shapiro L (1996) Cell cycle control by an essential bacterial two-component signal transduction protein. *Cell* **84:** 83-93

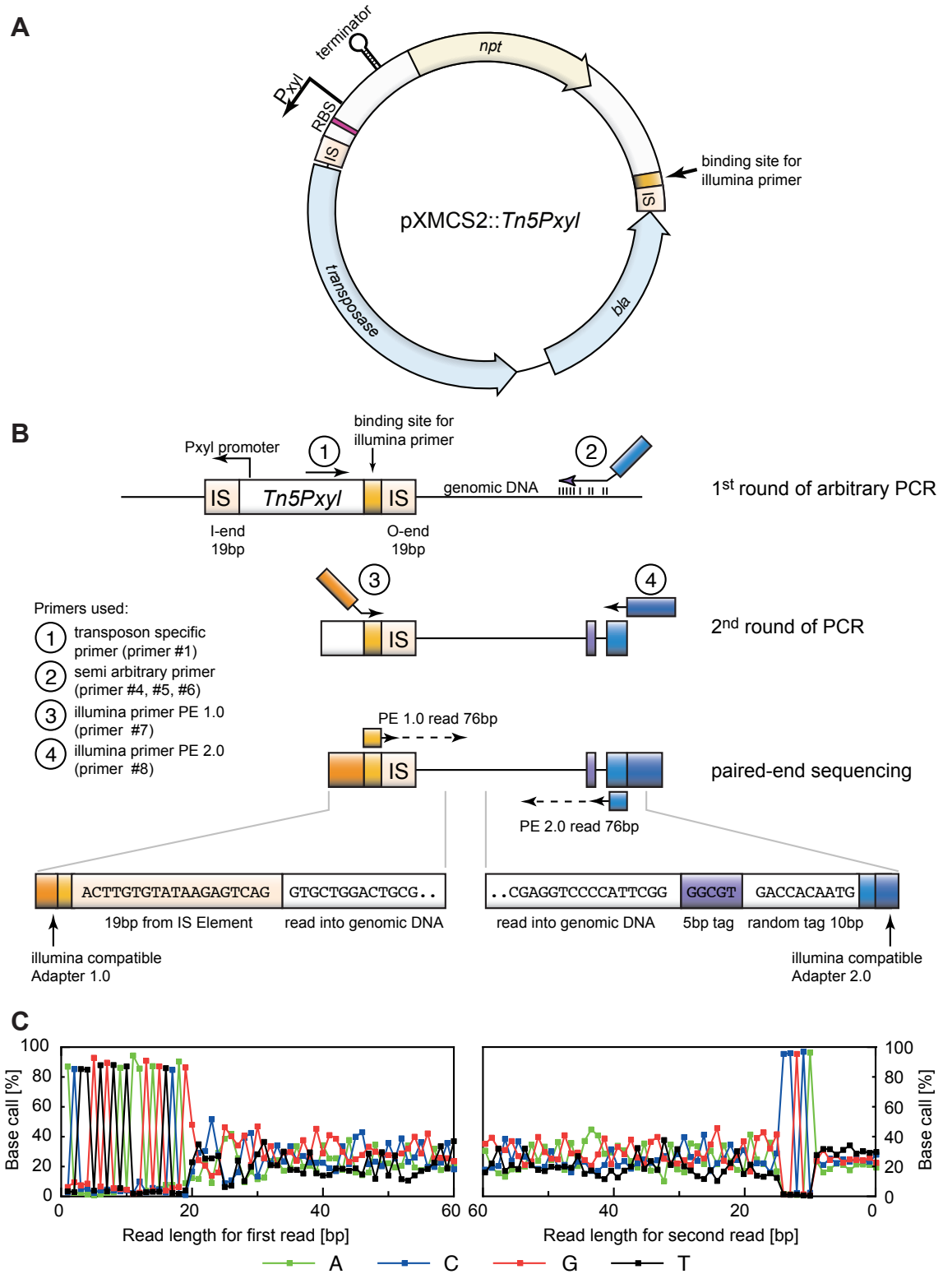
Stephens C, Reisenauer A, Wright R, Shapiro L (1996) A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **93:** 1210-1214

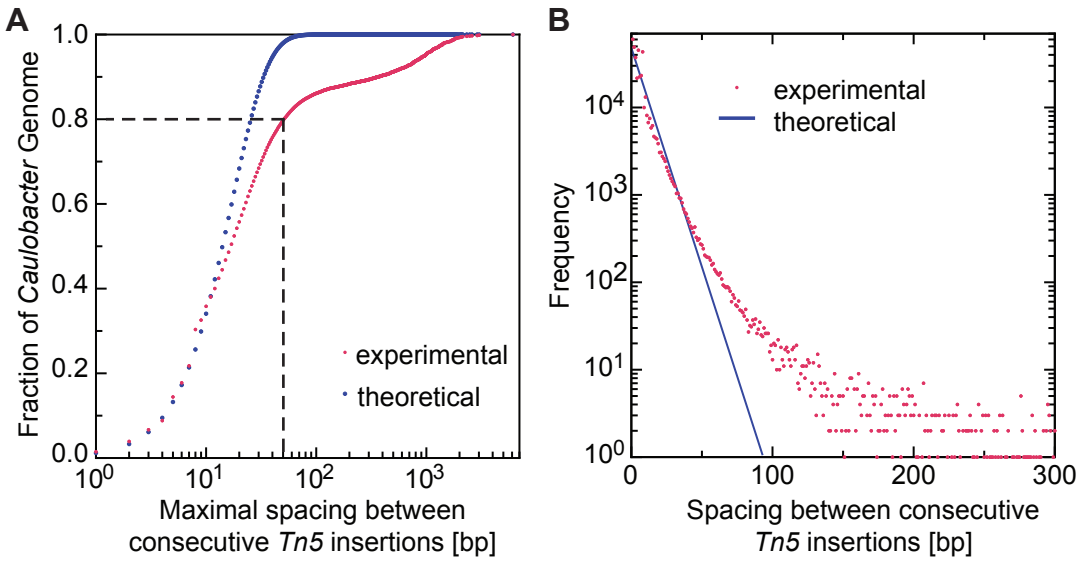
Tan MH, Kozdon JB, Shen X, Shapiro L, McAdams HH (2010) An essential transcription factor, SciP, enhances robustness of *Caulobacter* cell cycle regulation. *Proc Natl Acad Sci U S A* **107:** 18985-18990

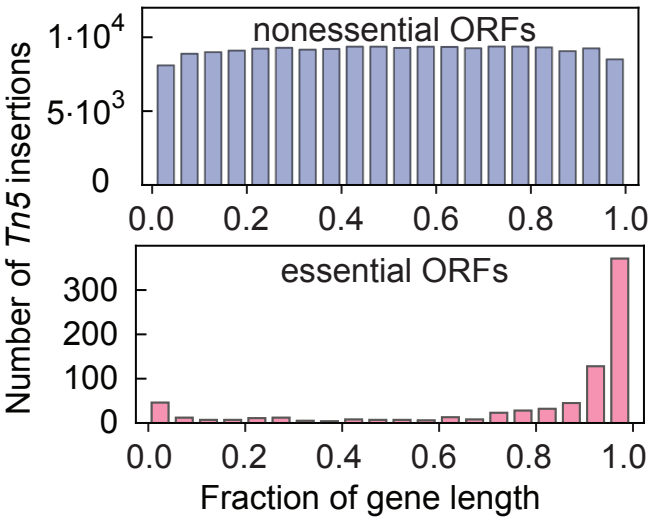
von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* **31:** 258-261

Wu J, Newton A (1997) The *Caulobacter* heat shock sigma factor gene rpoH is positively autoregulated from a sigma32-dependent promoter. *J Bacteriol* **179:** 514-521

Yamazaki Y, Niki H, Kato J (2008) Profiling of *Escherichia coli* Chromosome database. *Methods in molecular biology* **416:** 385-389


A
terminator
npt
PxyI
RBS
IS
binding site for
illumina primer
pXMCS2::Tn5PxyI
transposase
bla
B
PxyI promoter
binding site for
illumina primer
IS
Tn5PxyI
IS
genomic DNA
I-end
19bp
O-end
19bp
1st round of arbitrary PCR
Primers used:
1 transposon specific primer (primer #1)
2 semi arbitrary primer (primer #4, #5, #6)
3 illumina primer PE 1.0 (primer #7)
4 illumina primer PE 2.0 (primer #8)
2nd round of PCR
PE 1.0 read 76bp
paired-end sequencing
PE 2.0 read 76bp
ACTTGTGTATAAGAGTCAG
GTGCTGGACTGCG..
19bp from IS Element
read into genomic DNA
illumina compatible
Adapter 1.0
..CGAGGTCCCCATTCGG
GGCGT
GACCACAATG
read into genomic DNA
5bp tag
random tag 10bp
illumina compatible
Adapter 2.0
C
Base call [%]
Read length for first read [bp]
Read length for second read [bp]
A
C
G
T


A
1.0
0.8
0.6
0.4
0.2
0.0
Fraction of Caulobacter Genome
10^0
10^1
10^2
10^3
Maximal spacing between
consecutive Tn5 insertions [bp]
experimental
theoretical
B
10^4
10^3
10^2
10^1
10^0
Frequency
0
100
200
300
Spacing between consecutive
Tn5 insertions [bp]
experimental
theoretical


nonessential ORFs
1·10^4
5·10^3
0
0.0
0.2
0.4
0.6
0.8
1.0
essential ORFs
300
200
100
0
Number of Tn5 insertions
Fraction of gene length