SUPPLEMENTAL TEXT

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Mapping the transcription start site and identifying the region footprinted by the PhoP-P protein in PhoP-activated promoters. **(A)** S1 nuclease protection assay of the PhoP-activated *mig-14*, *pagK*, *yrbL*, *yobG*, *pagP*, *pipD*, *ybjX*, *slyB*, *pcgL*, *pagD*, *virK*, *ompX*, *mgtC* and *pgtE* promoters carried out with mRNAs harvested from wild-type (14028s) and *phoP* (MS7953s) *Salmonella* strains following growth in N-minimal medium with 10 mM (H) or 10 μM (L) Mg^{2+} concentrations. The arrow indicates the transcription start site. **(B)** DNase I footprinting analysis of promoter regions protected by the phosphorylated form of the PhoP protein (red lines) at the *mig-14*, *pagK*, *yrbL*, *yobG*, *pagP*, *pipD*, *ybjX*, *slyB*, *pcgL*, *pagD*, *virK*, *ompX*, *mgtC*, *rstA* and *pgtE* promoters. DNase I footprinting analysis has been previously reported for *mgtA* [2], *pagC* [2,3], *ugtL* [4], *phoP* [5] and *pmrD* [6].

Figure S2. DNA sequences of PhoP-activated promoters. DNA sequences corresponding to the 20 analyzed promoters. PhoP boxes are underlined and labeled by their corresponding identifier in Table S1. The central nucleotide of the PhoP box is highlighted (red color). The PhoP boxes in the reverse orientation are also presented by their reverse complementary sequence, which is the sequence format used to build the

PhoP submotifs. The RNAP -35 and -10 regions (blue color) and transcription start site (green color) are also indicated. These promoters are divided into five classes based on their shared architectures (see Fig. 4).

Figure S3. Engineered promoter constructs. Derivatives of the PhoP regulated *phoP*, *rstA*, *mgtC*, *pdgL*, *virK*, *yjbX* and *pagK* promoters. The PhoP boxes are indicated in lowercase, the RNAP -10 and -35 regions are underlined, and the transcription start site is highlighted in bold.

Figure S4. The α-CTD subunit of RNA polymerase is required to promote transcription of the *rstA* and *pagC* genes *in vitro*. Single-round *in vitro* transcription assays with linear templates corresponding to the *rstA* (**A**) and *pagC* (**B**) promoters, wild-type or mutant alpha Δ235 (α Δ-CTD) RNA polymerase, and increasing amounts of phosphorylated *Salmonella* PhoP-His-6 protein.

Figure S5. Models of interactions between a transcription factor and RNAP. **(A)** The CRP box, like the PhoP box, is functional when located at integral turns of the DNA helix with respect to the RNAP -10 hexamer, allowing upstream and/or downstream contacts with the α-CTD of RNAP [7]. **(B)** The CRP box is not functional when its face rotates 180˚ at half-integral turns of that hexamer [7], whereas **(C)** the PhoP box is functional at that face depending on its orientation and location with respect to the RNAP -10 hexamer. **(D)** The PhoP protein is predicted to contact the σ subunit of RNAP when the PhoP box completely overlaps the RNAP -35 hexamer. **(E)** Two-fold symmetry, and inverted repeat binding site of the CRP protein [7,9,10,11]. **(F)** Two-fold symmetry regulatory domain, and direct repeat binding site of the PhoP protein [12]. Head-to-tail DNA-binding domain (e.g., PhoB protein [8]) of the PhoP protein [12].

Figure S6. Classification of ancestral and horizontally acquired PhoP-activated genes. Hierarchical clustering (Statistical Toolbox, Matlab 2007b, complete linkage method and correlation similarity were applied) of the Conservation Scores (CS) of proteins encoded by 20 PhoP-activated genes in *Salmonella*. CS is calculated as in [13] (*e-value* <1E-5, expected number of false positives in a reciprocal BLAST search). CS values (represented by colors) can range from 55% to 100% when the closest homolog exhibits that percentage of amino acid identity. (Note that the phylogenetic relationships shown at the top of the figures do not represent phylogenetic distances.) **(A)** Group of genes exhibiting the highest CSs along the evaluated bacterial genomes. **(B)** Group of genes exhibiting the lowest CSs along the evaluated bacterial genomes. **(C)** Scatter plot combining the rank order derived from **(A)** with the CG-content (Bioinformatics Toolbox, Matlab 2007b) of the corresponding proteins in *Salmonella* (*F*-statistic *p-value* < 4.94e-05). Colors correspond to promoter architectures.

SUPPLEMENTAL TABLES

Table S1. Promoter dataset corresponding to 20 PhoP regulated promoters. ("Arch." follows the promoter description of Fig. 1; "PhoP box ID" designs the PhoP box in Fig. S1; "#" accounts for the number of PhoP box(es); "O" corresponds to the PhoP box orientation; "Function box" states for the role of the PhoP boxes, where "A" and "R" correspond to activation and repression; and "Distances" describes the distances between the PhoP boxes and the RNAP binding site;).

Gene Name	Function	Arch.	ID	Sequence_PhoP-box	#	O	Function PhoP box	Distances		
								boxes	-10	+1
yrbL	Putative cytoplasmic protein	I	s22	TCGTTTAGGTTTTGTTTAA	1	1	A		11	27
phoP	Response regulator	I	s8	TGGTTTATTAACTGTTTAT	1	1	A		12	33
mgtA	Mg2+ transporter	I	s33	TGGTTTATCGTTGGTTTAA	1	1	A		12	33
slyB	Outer membrane lipoprotein	I	s12	TCGTTTAAGATTGGTTAAT	1	1	A		12	32
pmrD	Connector of two-component systems	I	s10	CTATTGCCGTTTTGTTTAT	1	1	A		12	31
rstA	Response regulator	II	s11	TCGTTTAGAAAAGATTTAT	1	1	A		23	42
ompX	Outer membrane protein	II	s2	CGGTTGAGGGTTCGTTGAA	1	1	A		21	42
pcgL	D-ala-D-ala dipeptidase	III	s27	ATTTTAACCATCTGTTTAA	2	1	A		12	30
			s62	GAGTTTATATTTTGCTTAT		0	R	44		90
yobG	Putative inner membrane protein	III	s21	ACAGTTACTCCTGGTTTAA	2	1	A		12	31
			s70	GTTTTTAGGAATGATTCAC		0	R	15		62
pagP		III	s6	CTGTTTATAGTTTGTTAAG	2	1	A		23	43
			s73	TTTGTGAAAGCTTATTAAG		1	R	67		126
pagD		III	s74	TGGTTAACTCTTCGTTGAA	2	1	A		22	42
			s5	GTGTTTAGAGAGAATTTAC		1	R	86		144
virK		IV	s17	TCGTTGCCTTTACGTTTAA	2	1	A		12	32
			s15	CCATTGATAAACTGTTTAA		1	A	26		74
mig-14	Putative transcription activator	IV	s61	ACATTTTTATTTGGTTAAG	2	1	A		14	33
			s38	ATGTTTAGCTTGTATTTAA		0	A	8		57
ybjX		IV	s20	GTATTGACGATTGGTTAAT	2	1	A		12	31
			s18	TTGTTTAGATACGGTTTAC		0	A	30		77
ugtL	Inner membrane protein that modifies the LPS	V	s14	CGGTTGAGCAACTATTTAC	2	0	A		30	51
			s28	AATAATACTTTTAGTTTAA		0	R	-24		2

Gene Name	Function	Arch.	ID	Sequence_PhoP-box	#	O	Function PhoP box	Distances		
								boxes	-10	+1
pagK		V	s26	CCATTTATAAAATATTTAA	2	0	A		37	55
			s65	ACGTTTAATATCTATAGTA		1	R	-33		-2
pgtE	Outer membrane protease	V	s64	ATTTTTACCTTATATTGAA	2	0	A		49	68
			s7	ATGATTATAGATTGCTTAT		1	R	35		119
mgtC	Inner membrane virulence protein that aids grouth in low Mg2+	V	s35	CTGTTTAAGTTTGTTTGAT	2	0	A		48	67
			s34	ATGTTTAAACACGCTTTAT		1	R	-368		-301
			s36	ATGTTTCCTTATATTTTAA		1	R	-226		-159
pipD		V	s9	TTATTGAGGTTGTATTGAT	2	0	A		57	76
			s66	CCGTTACGCTGCTGGGTAT		1	R	-131		-55
pagC	Outer membrane protein	V	s4	GTGTTTAGAGAGAATTTAC	2	0	A		47	67
			s3	TTATTTACGGTGTGTTTAA		0	R	-594		-527

Table S2. Characterization of the RNAP binding sites corresponding to 20 PhoP regulated promoters (See Methods).

Gene Name	Score -10	Score -35	Distance -35_-10	Score RNAP	Extended Element
yrbL	0.67	0.33	18	0.62	TGTC
phoP	0.83	0.67	17	0.79	
mgtA	0.83	0.33	17	0.74	
slyB	0.83	0.50	14	0.63	TGAT
pmrD	0.50	0.33	15	0.54	TGTG
rstA	0.67	0.50	17	0.75	
ompX	0.83	0.83	19	0.79	
ybjX	0.67	0.33	18	0.61	GTGG
pdgL	0.50	0.33	20	0.54	
yobG	0.50	0.33	14	0.49	ATGA
pagP	0.83	0.33	16	0.69	TTGT
pagD	0.83	0.33	17	0.72	
virK	0.50	0.67	16	0.70	
mig-14	0.67	0.50	16	0.68	
ugtL	0.67	0.50	15	0.66	
pagK	0.67	0.50	15	0.65	
pgtE	0.50	0.50	17	0.68	TTGA
mgtC	1.00	0.67	16	0.91	TGAC
pipD	0.67	0.50	17	0.72	
pagC	0.83	0.33	17	0.71	

Table S3. Bacterial strains used in this study.

Strain or plasmid	Description	Ref. or source
S. enterica serovar Typhimurium		
14028s	wild-type	[14]
EG13918	*phoP*-HA	[15]
DH5α	F$^-$ *sup*E44 Δ*lac*U169 (ϕ80 *lac*ZΔM15) *hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1	[16]
Plasmids		
pMS201	pSC101*ori* Km^R, promoterless GFP*mut2* vector	[17]
pSGF101	pSC101*ori* Km^R, P*phoP*-GFP*mut2*	This work
pSGF102	pSC101*ori* Km^R, P*pmrD*-GFP*mut2*	This work
pSGF103	pSC101*ori* Km^R, P*slyB*-GFP*mut2*	This work
pSGF106	pSC101*ori* Km^R, P*ugtL*-GFP*mut2*	This work
pSGF107	pSC101*ori* Km^R, P*pagC*-GFP*mut2*	This work
pSGF109	pSC101*ori* Km^R, P*pagK*-GFP*mut2*	This work
pSGF111	pSC101*ori* Km^R, P*yobG*-GFP*mut2*	This work
pSGF112	pSC101*ori* Km^R, P*ugd*-GFP*mut2*	This work

Table S4. Promoter-specific primers used in DNaseI Footprinting and SI nuclease-protection assay

Number	Name	Sequence
6098	*A1: mig-14*	5' GCGGCATTAAACCGTCGCTTTCAC 3'
6099	*A2: mig-14*	5' GCCGTATTGCGTAAACACCTC 3'
9681	*A3: mig-14*	5' CTGTATCGCTACTGGATGGT 3'
9682	*A4: mig-14*	5' CCTCCCGGTATAGGGAAAAG 3'
9199	*B1: pagK*	5' GGATGGGAGCCAGTAAGGA 3'
9200	*B2: pagK*	5' CGCTCTTAACGTGTTTAATAATAC 3'
9220	*C1: yrbL*	5' TAAACCTCCCCCTCTATAC 3'
4519	*C2: yrbL*	5' GTCACGGTTGTAAATGAC 3'
9205	*D1: yobG*	5' CAGGGCAAACAATACAATTGCGC 3'
9204	*D2: yobG*	5' CGATGCCGAGAACGACCCATC 3'
5598	*E1: pagP*	5' AAATCACGGCGCGGTTATTTTTTCATGTC 3'
6527	*E2: pagP*	5' GAAGGAATAGCCAGCCATGG 3'
6102	*F1: slyB*	5' GGCCCGATTTCATACCTCTTCC 3'
6103	*F2: slyB*	5' ACCCCATCAGTGAAACGGCCAG 3'
9206	*G1: pdgL*	5' CTGTACTACTGATTATCAAATG 3'
9207	*G2: pdgL*	5' CAGGCTGAACAGGAGGAAAG 3'
7981	*H1: pagD*	5' CCCGCTCCGATTTAATCT 3'
7982	*H2: pagD*	5' CTGGCCAAAACATGGAATG 3'
8201	*I1: virK*	5' CGCTTGCCGCCTTCCTGCAACT 3'
8202	*I2: virK*	5' CGATAAGGCGTCTGCGCCATG 3'
9683	*I3: virK*	5' CTTGCCGCCTTCCTGCAACT 3'
9684	*I4: virK*	5' CATATCACTTTGCTGCATC 3'
9297	*J1: ompX*	5' GTCAGTACGAGTAAAGGTGGCA 3'
9298	*J2: ompX*	5' CCAGTGCTGAAAGACATGCA 3'
8203	*K1: pipD*	5' CGCTATGCCGATGAACATAATATC 3'
8204	*K2: pipD*	5' GCCATCTTCGTTGCGCGCGA 3'
8431	*L1: ybjX*	5' ATCATTCGCCAGCCAGCCATC 3'
8432	*L2: ybjX*	5' CGAAAGCTGGCTTTACGCCAG 3'
6891	*M1: mgtC*	5' AATTCATGCAGGAGTAATATG 3'
1250	*M2: mgtC*	5' GAATTCTGGAAGAATAAGTACGTGC 3'
9732	*N1: rstA*	5' AGAGGAATTAATCCGGCGA 3'
9733	*N2: rstA*	5' ACGCGGCTCAACAATGACATC 3'

Table S5. Promoter-specific primers used in SI nuclease-protection assay

Number	Name	Sequence
4518	*C3: yrbL*	5' GATGACAGTCACCCCAAAACC 3'
4519	*C2: yrbL*	5' GGTCACGGTTGTAAATGAC 3'
9198	*J3: ompX*	5' CGGCGGTTGAGGGTTCGTTGA 3'
9202	*J4: ompX*	5' CCAGAACAGCGGCCAGTGCTG 3'
8203	*K1: pipD*	5' CCGCTATGCCGATGAACATAATATC 3'
9575	*K3: pipD*	5' GCGATGGGATGAATAACCTTATG 3'
6005	*M3: mgtC*	5' GGAATTTATTGTTTAATGATTTCAGACG 3'
6002	*M4: mgtC*	5' CGCGTGTTGTAACATTATTGGGATG 3'

Table S6. Primers used to construct the promoters in GFP reporter plasmids

Number	Name	Sequence
4811	*phoP*	forward 5′ CGGCTCGAG AAGAGTTGACCCGTGGCAAGCGTG 3′
4432	*phoP*	reverse 5′ CGGGATCC TGGCGTAATAATGCATTATCCTC 3′
4842	*rstA*	forward 5′ CGGCTCGAG AAATAAGGAATGATCGACCAC 3′
5231	*rstA*	reverse 5′ CGGGATCC CAGATAAGCGGCAATGAGAG 3′
8160	*rstA3*	forward 5′ CGGCTCGAG CGGAACCGCTCTCGTTTAGA 3′
5231	*rstA*	reverse 5′ CGGGATCC CAGATAAGCGGCAATGAGAG 3′
5228	*pagK*	forward 5′ CGGCTCGAG AACAGCAGAAGAGTTGTTATG 3′
5227	*pagK*	reverse 5′ CGGGATCC GAAGATGGTAATATTAAAACC 3′
4737	*mgtC*	forward 5′ CGGCTCGAG GAGCTATCGCCGGTATTAAGCAGG 3′
4736	*mgtC*	reverse 5′ CGGGATCC CATGAACATGACGATTACCT 3′
8142	*phoP4*	forward 5′ TCGA CTATTTGTCTGGTTTATTAACTGTTTATCCCCAAAGCACCATAATCAACGCTAGACTG 3′
8143	*phoP4*	reverse 5′ GATC CAGTCTAGCGTTGATTATGGTGCTTTGGGGATAAACAGTTAATAAACCAGACAAATAG 3′
8144	*rstA4*	forward 5′ TCGA AACCGCTCTCGTTTAGAAAAGATTTATGGAAGGGAAAACAGAGGCGGTGTATGTTGGCGTTTTCTA 3′
8145	*rstA4*	reverse 5′ GATC TAGAAAACGCCAACATACACCGCCTCTGTTTTCCCTTCCATAAATCTTTTCTAAACGAGAGCGGTT 3′
8146	*pagK4*	forward 5′ TCGA TTATTTAAGTTAAATATTTTATAAATGGTTTTTATTTACTCACCTGATGGTAATGAATAACGTTTAATATCTATAGTAAAGG 3′
8147	*pagK4*	reverse 5′ GATC CCTTTACTATAGATATTAAACGTTATTCATTACCATCAGGTGAGTAAATAAAAACCATTTATAAAATATTTAACTTAAATAA 3′
8148	*mgtC4*	forward 5′ TCGA ATCAAACAAACTTAAACAGAACGTCACTAAACCCGCCTTTGCACTTTACGGAACATATTGGCTGACTATAATAAGCGCAAATT 3′
8149	*mgtC4*	reverse 5′ GATC AATTTGCGCTTATTATAGTCAGCCAATATGTTCCGTAAAGTGCAAAGGCGGGTTTAGTGACGTTCTGTTTAAGTTTGTTTGAT 3′
8150	*phoP4-rev*	forward 5′ TCGA CTATTTGTCATAAACAGTTAATAAACCACCCCAAAGCACCATAATCAACGCTAGACTG 3′
8155	*phoP4-rev*	reverse 5′ GATC CAGTCTAGCGTTGATTATGGTGCTTTGGGGTGGTTTATTAACTGTTTATGACAAATAG 3′
8162	*rstA4-rev*	forward 5′ TCGA AACCGCTCATAAATCTTTTCTAAACGAGGAAGGGAAAACAGAGGCGGTGTATGTTGGCGTTTTCTA 3′
8165	*rstA4-rev*	reverse 5′ GATC TAGAAAACGCCAACATACACCGCCTCTGTTTTCCCTTCCTCGTTTAGAAAAGATTTATGAGCGGTT 3′

Table S7: Genome-wide analysis of PhoP activated genes in *Klebsiella pneumoniae* using gene expression measured by microarray experiments of wild-type (percentage of gene covered by expressed probe pairs; averaged and maximum values of expressed probe pairs) and phoP mutated strains (Nimblegen tiling arrays), and promoter binding by the PhoP protein, measured by a ChIP-chip assay and identified by using the D&C method [1], where more than one PhoP box can be protected under the same ChIP peak (Position is relative to the upstream end of the corresponding promoter region; C indicates reverse orientation; Score is defined in [0,1]).

Gene			**Expression**			**ChIP**	**1st Submotif**			**2nd Submotif**		
Code	**Name**	**Orient.**	**Coverage %**	**Average**	**Maximum**	**Position**	**Position**	**Submotif**	**Score**	**Position**	**Submotif**	**Score**
KPN_04773	*N/A*	-	100	3.79	5.26	126	209C	3	0.97	182	3	0.7
KPN_03847	*yfbE*	-	100	10.75	22.35	8	181	5	0.86	79	1	0.85
KPN_02444	*pmrD*	+	100	5.92	8.61	61	176C	4	0.77			
KPN_01136	*phoP*	-	100	13.71	29.55	100	209	8	0.72	220	1	0.56
KPN_01979	*slyB*	+	100	2.43	2.88	100	125	2	0.72	82C	12	0.56
KPN_02997	*N/A*	-	100	3.24	3.80	53	287C	1	0.65	298C	11	0.56
KPN_00873	*ybjG*	-	100	5.11	11.01	84	134	7	0.51			
KPN_02355	*N/A*	+	98	4.19	9.87	22	76C	10	0.76	65C	3	0.75
KPN_01508	*N/A*	+	96	2.60	3.25	143	125	9	0.59			
KPN_02050	*N/A*	-	95	5.78	8.59	100	194C	12	0.71			
KPN_01450	*N/A*	-	92	4.45	10.23	5	138	8	0.7			
KPN_04654	*mgtA*	+	79	3.12	11.85	312	70	4	0.88			
KPN_02341	*yebO*	-	63	1.84	5.16	283	93	10	0.75			
KPN_03285	*fimB*	+	35	0.89	4.16	420	126	6	0.62	513	8	0.52
KPN_00559	*N/A*	+	26	0.58	2.79	346	9	2	1			
KPN_02051	*N/A*	+	17	0.44	3.00	223	249	12	0.71			
KPN_01847	*sfcA*	+	8	0.17	2.17	172	62	11	0.57			
KPN_01521	*rstA*	-	7	0.16	2.25	93	11C	11	0.83			
KPN_01602	*ygbM*	-	7	0.17	2.53	16	24	10	0.54			
KPN_01492	*N/A*	-	4	0.09	2.30	49	160	1	0.66			
KPN_02399	*N/A*	+	4	0.09	2.48	124	278	6	0.58	60C	5	0.54
KPN_02443	*N/A*	-	3	0.10	3.13	123	79	4	0.77			
KPN_03124	*pyrG*	-	3	0.06	2.11	236	20	7	0.5			
KPN_00564	*N/A*	+	3	0.06	2.15	148	28	5	0.51			
KPN_00843	*ompX*	+	1	0.01	2.42	252	65	12	0.86	297	5	0.6

REFERENCES

1. Harari O, Park SY, Huang H, Groisman EA, Zwir I (2010) Defining the plasticity of transcription factor binding sites by Deconstructing DNA consensus sequences: the PhoP-binding sites among gamma/enterobacteria. PLoS Comput Biol 6: e1000862.
2. Lejona S, Aguirre A, Cabeza ML, Garcia Vescovi E, Soncini FC (2003) Molecular characterization of the Mg2+-responsive PhoP-PhoQ regulon in Salmonella enterica. J Bacteriol 185: 6287-6294.
3. Perez JC, Latifi T, Groisman EA (2008) Overcoming H-NS-mediated transcriptional silencing of horizontally acquired genes by the PhoP and SlyA proteins in Salmonella enterica. J Biol Chem 283: 10773-10783.
4. Shi Y, Latifi T, Cromie MJ, Groisman EA (2004) Transcriptional control of the antimicrobial peptide resistance ugtL gene by the Salmonella PhoP and SlyA regulatory proteins. J Biol Chem 279: 38618-38625.
5. Soncini FC, Vescovi EG, Groisman EA (1995) Transcriptional autoregulation of the Salmonella typhimurium phoPQ operon. J Bacteriol 177: 4364-4371.
6. Kato A, Latifi T, Groisman EA (2003) Closing the loop: the PmrA/PmrB two-component system negatively controls expression of its posttranscriptional activator PmrD. Proc Natl Acad Sci U S A 100: 4706-4711.

7. Gaston K, Bell A, Kolb A, Buc H, Busby S (1990) Stringent spacing requirements for transcription activation by CRP. Cell 62: 733-743.

8. Blanco AG, Sola M, Gomis-Ruth FX, Coll M (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. Structure 10: 701-713.

9. Pedersen H, Valentin-Hansen P (1997) Protein-induced fit: the CRP activator protein changes sequence-specific DNA recognition by the CytR repressor, a highly flexible LacI member. Embo J 16: 2108-2118.

10. Rhodius VA, West DM, Webster CL, Busby SJ, Savery NJ (1997) Transcription activation at class II CRP-dependent promoters: the role of different activating regions. Nucleic Acids Res 25: 326-332.

11. Weber IT, Steitz TA (1987) Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 A resolution. J Mol Biol 198: 311-326.

12. Bachhawat P, Stock AM (2007) Crystal structures of the receiver domain of the response regulator PhoP from Escherichia coli in the absence and presence of the phosphoryl analog beryllofluoride. J Bacteriol 189: 5987-5995.

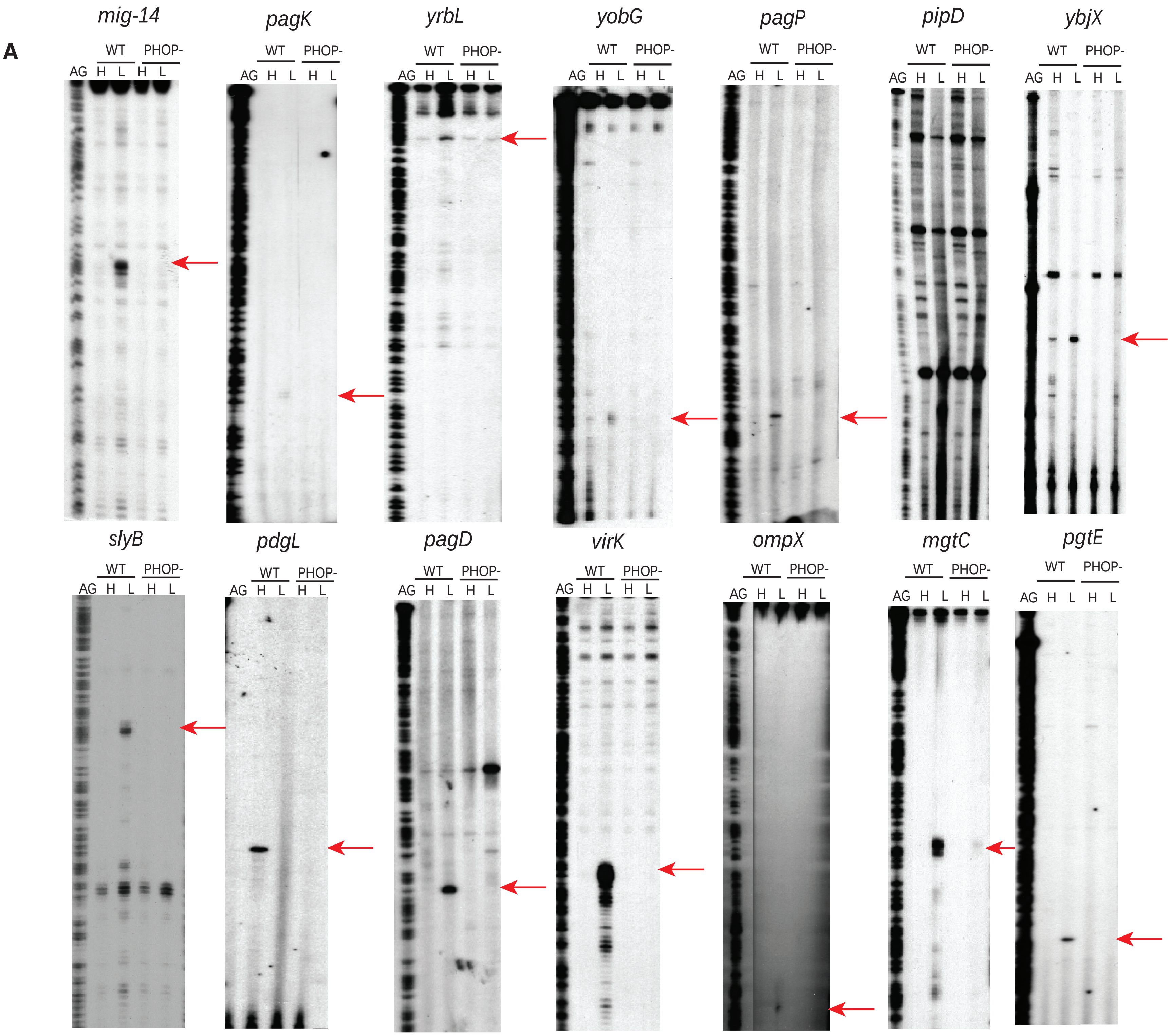
13. Janky R, van Helden J (2007) Discovery of conserved motifs in promoters of orthologous genes in prokaryotes. Methods Mol Biol 395: 293-308.

14. Fields PI, Swanson RV, Haidaris CG, Heffron F (1986) Mutants of Salmonella typhimurium that cannot survive within the macrophage are avirulent. Proc Natl Acad Sci U S A 83: 5189-5193.

15. Shin D, Groisman EA (2005) Signal-dependent Binding of the Response Regulators PhoP and PmrA to Their Target Promoters in Vivo. J Biol Chem 280: 4089-4094.

16. Hanahan D (1983) Studies on transformation of Escherichia coli with plasmids. J Mol Biol 166: 557-580.

17. Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. Proc Natl Acad Sci U S A 100: 11980-11985.


A
mig-14
pagK
yrbL
yobG
pagP
pipD
ybjX
slyB
pdgL
pagD
virK
ompX
mgtC
pgtE
WT
PHOP-
AG
H
L
H
L

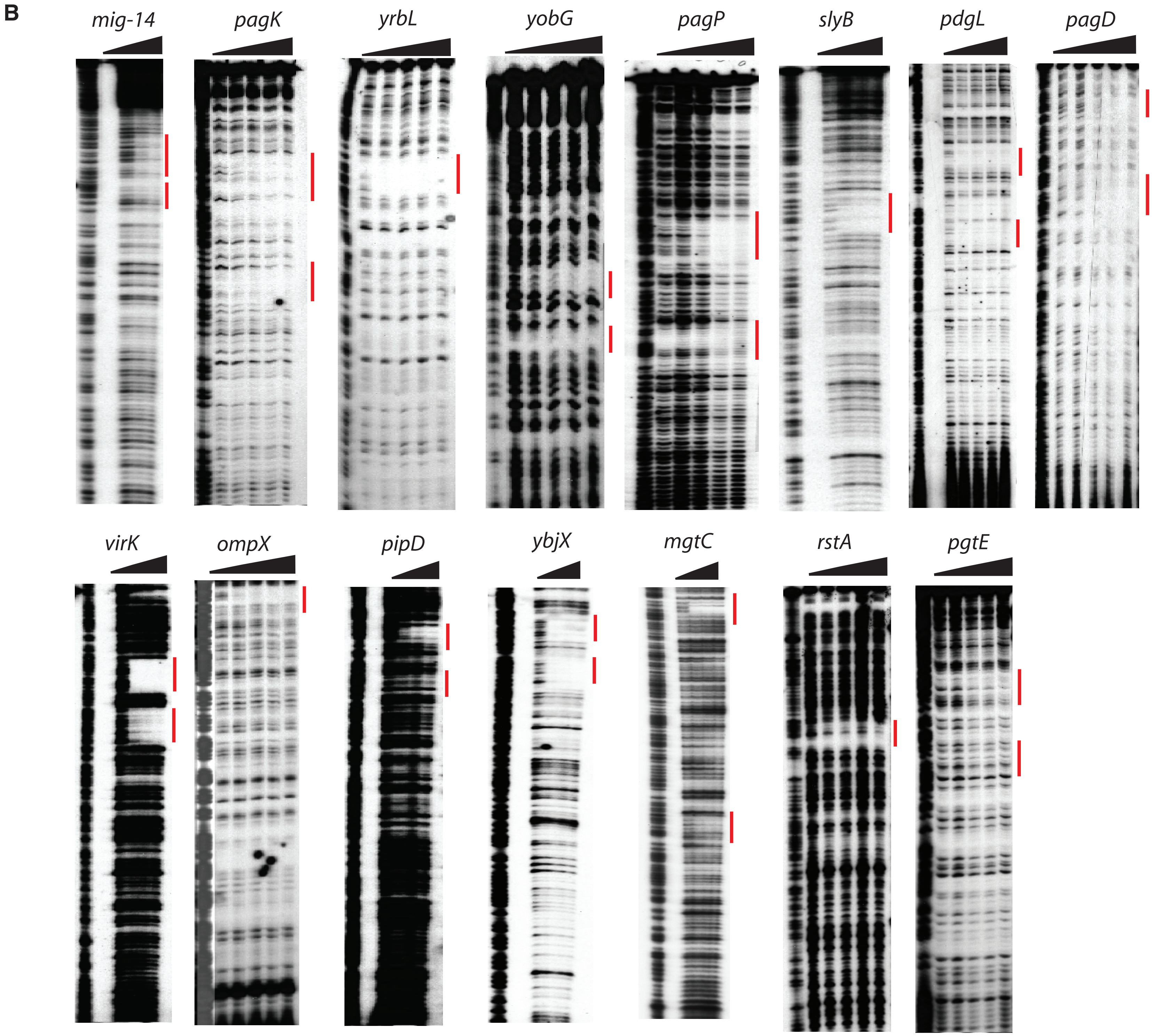

B
mig-14
pagK
yrbL
yobG
pagP
slyB
pdgL
pagD
virK
ompX
pipD
ybjX
mgtC
rstA
pgtE

Figure S1

```
      100-------90--------80--------70--------60--------50--------40--------30--------20--------10--------0

mgtA     GATGGTGATAAAAAGAGTAACTTTACTCAAGCGAAGGGCAATGGGCAAAGTCTGGTTTATCGTTGGTTTAATTACGTAACGGTATGATACCGCCATAAT
                                                                      s33
phoP     AATGTGTGCCTGCCTCACCCTCTTTTCTTCAGAAAGAGGGTGACTATTTGTCTGGTTTATTAACTGTTTATCCCCAAAGCACCATAATCAACGCTAGAC
                                                                      s8
yrbL     CATTTCGTGCGCTTTACCTATGCATAAAAACATTAAGAAAACCTTAAAAAATTTCGTTTAGGTTTTGTTTAAGTTCTTTGTCCATACTAAATTTATATG
                                                                      s22
slyB     TTTGCCGTGTCGTCCACGGATATCAATGACCTATCTTTTCTTGCTTCCGACTTCGTTTAAGATTGGTTAATTAACTTCTGATTATGATTTTCACATTCT
                                                                     s12
pmrD     TGTGTAAGATCTGTAAAAATTAATGTTAGGTTAATATACAACCATTCCATCGCTATTGCCGTTTTGTTTATCCGTTGATGTGTAACGTGTTATGAACAA
                                                                     s10

rstA     ACACGCTAATAACAACAGCGTGTAGCATAACGGAACCGCTCTCGTTTAGAAAAGATTTATGGAAGGGAAAACAGAGGCGGTGTATGTTGGCGTTTTCTA
                                                          s11
ompX     ACAGTACGAGTAAAGGTGGCAGTGTAATGGAAATTCCCCGCGGCGGTTGAGGGTTCGTTGAAAAAAAATTTCCAGGATTCCATAAAATGAACTTAAATT
                                                            s2

yobG     GTGTACATTGATTGAAACAAGTGAATCATTCCTAAAAACATATAATGCCGCTACAGTTACTCCTGGTTTAAGTAACTCATGATAATTTAGCGACATAAG
                             GTTTTTAG---TGATTCAC                     s21

pcgL     ATATAAACTCTCCGTTTGTGATAAGGCACAGATTACAGGGGGAAAGATGTTTATTTTAACCATCTGTTTAAGCGGTGTCCGCTAATCTTTACCTGCTCA
         <-ATAAGCAAAAT                                                s27
           GAGTTTAT---TTGCTTAT s62

pagP     AAATCAAACGCCGTTAACCCGATACTCTCTCAGATTATTCTCTGTTTATAGTTTGTTAAGATTTTATTCAGGTTAATGTTGTTATTATCACAGTCGAAT
         <-25 TTTGTGAAAGCTTATTAAG                         s6

pagD     TAAAAAAGTCTATTTCGCTATAAAACCCATTATTATTAAGAGTGGTTAACTCTTCGTTGAATAAAAAATGTCAATGACGTTCCATAATTCAGGAGATGA
         <-42 GTGTTTAGAGAGAATTTAC s5

virK     TTATTACCGCCATTGATAAACTGTTTAACAACATCGTCTGTACAGACCTTCTTCGTTGCCTTTACGTTTAACTCAATCAGGCTACCGTCTCGGTTATAA
                          s15                                        s17

mig-14   AACGTCATCGATTAGCATGTTAACCATTAAATACAAGCTAAACATTTGTCACATTTTTATTTGGTTAAGCAAAAAAATAATACAAAATAGCATTTTCAG
                                   ATGTTTAG---GTATTTAA s38         s61

ybjX     AGAAAGTAAACCGTATCTAAACAAACCTGACGCAATAAAGATAGCGCCGGCTGTATTGACGATTGGTTAATGTTTCCTGTGGTTCAATCTCGACAATCT
              TTGTTTAG---CGGTTTAC s18                                s20

pagK     TAATAGTGCTTATCATTTTTATTTAAGTTAAATATTTTATAAATGGTTTTTATTTACTCACCTGATGGTAATGAATAACGTTTAATATCTATAGTAAAG
                                    CCATTTAT---ATATTTAA s26                                   s65

ugtL     TGATTAATTACGCACTATATTATTTTTAGAGAAAGTAAATAGTTGCTCAACCGTGTAGAAATTGTCTTATAAGAAGTTAAACTAAAAGTATTATTAGGC
                                           CGGTTGAG---CTATTTAC s14                          s28

pgtE     ATTCTATTTCTTAACTTCAATATAAGGTAAAAATGCGTCAAGTTCTCTGGCGTAATAAATGTACTCTTGTCCGACGATTTGACAAGATGAAAACTTCAT
                        ATTTTTAC---ATATTGAA s64
         <-18 ATGATTATAGATTGCTTAT s7

mgtC     TGTCATAAGAAAAAATATCAAACAAACTTAAACAGAACGTCACTAAACCCGCCTTTGCACTTTACGGAACATATTGGCTGACTATAATAAGCGCAAATT
                          CTGTTTA---ATGTTTGAT s35
         172-> ATGTTTAAACACGCTTTAT s34 106-> ATGTTTCCTTATATTTTAA s36

pagC     TATTATTTTTTTGGAATGTAAATTCTCTCTAAACACAGGTGATATTTATGTTGGAATTGTGGTGTTGATTCTATTCTTATAATATAACAAGAAATGTTG
                          GTGTTTAG---GAATTTAC s4
         513-> TTAAACACACCGTAAATAA
                TATTTAC---GTGTTTA s3

pipD     TGTGGTTATCAATACAACCTCAATAAAGAGAAAATAAAGGATGCCGGCAAAGAATATTGCTGGCTCGATCACTTTCTCGGTTTCCAATCCAGGGTCAAT
                TTATTGAG---GTATTGAT s9
         ->56 CCGTTACGCTGCTGGGTAT s66
```

Figure S2

A

phoP4 TCGACTATTTGTCTggtttattaactgtttatCCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
rstA4 TCGAAACCGCTCtcgtttagaaaagatttatGGAAGGGAAAACAGAGGCGGTGTATGTTGGCGT**T**TTCTAGATCC

phoP AAGAGTTGACCCGTGGCAAGCGTGTTGATGCCGAAGGAATGAAACAGTTTATTGATAGTCTGGCCCTGCCGGAAGCAGAAAAAACGCGCCTTAAAGCCATGACGCCGGCAAATTAT
ATCGGTCGCGCTGTGACTCTGGTCGACGAACTTAAATAATGCCTGCCTCACCCTCTTTTCTTCAGAAAGAGGGTGACTATTTGTCT*ggtttattaactgttta*TCCCCAAAGCACC
ATAATCAACGCT**A**GACTGTTCTTATTGTTAACACAAGGGAGAAGAG**ATG**ATGCGCGTACTGGTTGTAGAGGATAATGCATTATTACGCCA

rstA AAATAAGGAATGATCGACCACATGCTAAAGGCGATGGTGGTACGCATCGCGTCGATGCCCCGCTCGCTGGCCACGATATAGTGCGCGATAAGCGCAAACGTTGGAAACAGAGGAA
TTAATCCGGCGATATAATAATTTTTCGTTTTTGACAGCAGACCAATCAATACGACGACCAGCGCCCCCAGGGCAGCTTTAATCACGAGTCCCATTACCTTGCCTTAACACGCTAA
TAACAACAGCGTGTAGCATAACGGAACCGCTCTcgtttagaaaagatttaTGGAAGGGAAAACAGAGGCGGTGTATGTTGGCGT**T**TTCTATTCTCCATTTATAAT**ATG**AACCGCA
TTGTATTTGTTGAAGACGACGCTGAAGTGGGTTCTCTCATTGCCGCTTATCTG

ybjX4 TCGAGCAAAATTTTAACCAGAAAgtaaaccgtatctaaacaaACCTGACGCAATAAAGATAGCGCCGGCTgtattgacgattggttaatGTTTCCTGTGGTTCAATCTCGA**C**AATCT
ybjX4-ΔUP TCGAGCAAAATTTTAACCAGAAAGTAAACCGTATCTtAtGAAACCTGACGCAATAAAGATAGCGCCGGCTgtattgacgattggttaatGTTTCCTGTGGTTCAATCTCGA**C**AATCT
ybjX4-ΔDN TCGAGCAAAATTTTAACCAGAAAGTAAACCGTATCTAAACAAACCTGACGCAATAAAGATAGCGCCGGCTgtattgacgattggatatGTTTCCTGTGGTTCAATCTCGA**C**AATCT

virK4 TCGATTTATTACCGCCATTgataaactgtttaacaacaTCGTCTGTACAGACCTTCTtcgttgcctttacgtttaaCTCAATCAGGCTACCGTCTCGGTT**A**TAAGT
virK4-ΔUP TCGATTTATTACCGccattgataaactgaataaCAACATCGTCTGTACAGACCTTCTtcgttgcctttacgtttaaCTCAATCAGGCTACCGTCTCGGTT**A**TAAGT
virK4-ΔDN TCGATTTATTACCGCCATTGATAAACTGTTTAACAACATCGTCTGTACAGACCTTCTtcgttgcctttacgaataaCTCAATCAGGCTACCGTCTCGGTT**A**TAAGT

mig144 TCGATGTTAACCAttaaatacaagctaaacatTTGTCAcatttttatttggttaaGCAAAAAAATAATACAAAATAGCAT**T**TTCAG
mig144-ΔUP TCGATGTTAACCAttaaatacaagctattcatTTGTCAcatttttatttggttaaGCAAAAAAATAATACAAAATAGCAT**T**TTCAG
mig144-ΔDN TCGATGTTAACCAttaaatacaagctaaacatTTGTCAcatttttatttccttaaGCAAAAAAATAATACAAAATAGCAT**T**TTCAG

pcgL4-wt AGTGGTCGTAtaagcaaaatataaactCTCCGTTTGTGATAAGGCACAGATTACAGGGGGAAAGATGTTTAttttaaccatctgtttaAGCGGTGTCCGCTAATCTTTAC**C**TGCTC
pcgL4-ΔUP AGTGGTCGTAtaagcaaaatatattctCTCCGTTTGTGATAAGGCACAGATTACAGGGGGAAAGATGTTTAttttaaccatctgtttaAGCGGTGTCCGCTAATCTTTAC**C**TGCTC
pcgL4-ΔDN AGTGGTCGTAtaagcaaaatataaactCTCCGTTTGTGATAAGGCACAGATTACAGGGGGAAAGATGTTTAttttaaccatctgaataAGCGGTGTCCGCTAATCTTTAC**C**TGCTC

mgtC4 TCGAatcaaacaaacttaaacagAACGTCACTAAACCCGCCTTTGCACTTTACGGAACATATTGGCTGACTATAATAAGCG**C**AAATTGATCC

mgtC GAGCTATCGCCGGTATTAAGCAGGAATTTATTGTTTAATGATTTCAGACGAGCTTGTTATCGACATAATATTGTCATTTTTTTGTCATAAGAAAAAATATcaaacaa**a**cttaaa
cagAACGTCACTAAACCCGCCTTTGCACTTTACGGAACATATTGGCTGACTATAATAAGCG**C**AAATTCATGCAGGAGTAATATGTTGGACAGTCACTTTTACGTAAATCATCTG
GCAAGTTAACGCACGCTATTCCTGCGCTGCTTGCCGAACCGGTGGGCAGCAATCTCCCCTTGTGACGATTGTCATCCCAATAATGTTACAACACGCGCATTGTCGCGAGGTAAT
CGTCATGTTCAtgtttaaacacgctttaTTTCCTCCGCCGTTAACACGACGCTAATTGCCTCAGGGCAGAAATTTGTCGTGTGCTAAATATAGCACGTACTTATTCTTCCAGAA
AAA**ATG**GAGGAACGTATGTTAAtgtttccttatattttaAATTTACTGGCCGCT

phoP4/2 TCGATTTATTACCGCCATTgataaactgtttaacaacaTCGTCTGTACAGACCTTCTtggtttattaactgtttatCCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC

B

phoP4	TCGACTATTTGTCtggtttattaactgtttatCCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev	TCGACTATTTGTCataaacagttaataaaccaCCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
rstA4	TCGAAACCGCTCtcgtttagaaaagatttatGGAAGGGAAAACAGAGGCGGTGTATGTTGGCGT**T**TTCTAGATCC
rstA4-rev	TCGAAACCGCTCataaatcttttctaaacgaGGAAGGGAAAACAGAGGCGGTGTATGTTGGCGT**T**TTCTAGATCC

phoP4	TCGACTATTTGTCtggtttattaactgtttatCCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4+9nt	TCGACTATTTGTCtggtttattaactgtttatAAGGGAAAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4+10nt	TCGACTATTTGTCtggtttattaactgtttatGAAGGGAAAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4+11nt	TCGACTATTTGTCtggtttattaactgtttatGGAAGGGAAAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4+12nt	TCGACTATTTGTCtggtttattaactgtttatTGGAAGGGAAAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4+13t	TCGACTATTTGTCtggtttattaactgtttatCTGGAAGGGAAAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4+14nt	TCGACTATTTGTCtggtttattaactgtttatTCTGGAAGGGAAAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4+15nt	TCGACTATTTGTCtggtttattaactgtttatCTCTGGAAGGGAAAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4+16nt	TCGACTATTTGTCtggtttattaactgtttatGCTCTGGAAGGGAAAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4+21nt	TCGACTATTTGTCtggtttattaactgtttatGAACCGCTCTGGAAGGGAAAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC

phoP4-revK’	TCGAGAAAGAGGataaacagttaataaaccaGTGACTATTTGTCTGGTTTATTAACTTTGACAGTTTATCCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-revK’-dir	TCGAGAAAGAGGtggtttattaactgtttatGTGACTATTTGTCTGGTTTATTAACTTTGACAGTTTATCCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC

phoP4-rev-C	TCGACTATTTGTCataaacagttaataaaccaAACGTCACTAAACCCGCCTTTGCACTTTACGGTTATCCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K	TCGACTATTTGTCataaacagttaataaaccaTTTTTATTTACTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC

phoP4-rev-K	TCGACTATTTGTCataaacagttaataaaccaTTTTTATTTACTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K-5nt	TCGACTATTTGTCataaacagttaataaaccaATTTACTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K-6nt	TCGACTATTTGTCataaacagttaataaaccaTTTACTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K-7nt	TCGACTATTTGTCataaacagttaataaaccaTTACTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K-8nt	TCGACTATTTGTCataaacagttaataaaccaTACTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K-9nt	TCGACTATTTGTCataaacagttaataaaccaACTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K-10nt	TCGACTATTTGTCataaacagttaataaaccaCTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K-11nt	TCGACTATTTGTCataaacagttaataaaccaTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K+10nt	TCGACTATTTGTCataaacagttaataaaccaTTATACTTGCTTTTTATTTACTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K+11nt	TCGACTATTTGTCataaacagttaataaaccaATTATACTTGCTTTTTATTTACTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC

C

phoP4+11nt	TCGACTATTTGTCtggtttattaactgtttatGGAAGGGAAAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4+11nt-10	TCGACTATTTGTCtggtttattaactgtttatGGAAGGGAAAACCCCAAAGCACTATGTTCAACGCT**A**GACTGGATCC
phoP4-rev-C	TCGACTATTTGTCataaacagttaataaaccaAACGTCACTAAACCCGCCTTTGCACTTTACGGTTATCCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-C-10	TCGACTATTTGTCataaacagttaataaaccaAACGTCACTAAACCCGCCTTTGCACTTTACGGTTATCCCCAAAGCACTATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K	TCGACTATTTGTCataaacagttaataaaccaTTTTTATTTACTCACCTGATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
phoP4-rev-K-35	TCGACTATTTGTCataaacagttaataaaccaTTTTTATTTACTCTTGACATGGTAACCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC
mgtC4-P-box	TCGAataaacagttaataaaccaAACGTCACTAAACCCGCCTTTGCACTTTACGGAACATATTGGCTGACTATAATAAGCG**C**AAATTGATCC
phoP4-rev-C	TCGACTATTTGTCataaacagttaataaaccaAACGTCACTAAACCCGCCTTTGCACTTTACGGTTATCCCCAAAGCACCATAATCAACGCT**A**GACTGGATCC

Figure S3

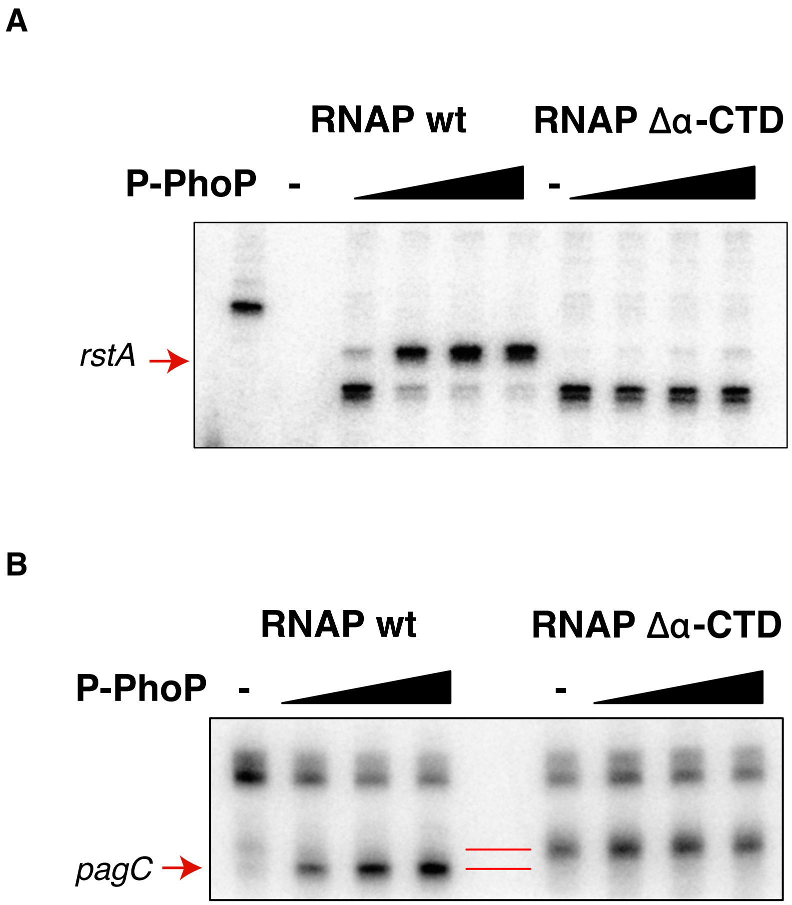

A
RNAP wt
RNAP Δα-CTD
P-PhoP
-
-
rstA
B
RNAP wt
RNAP Δα-CTD
P-PhoP
-
-
pagC

Figure S4

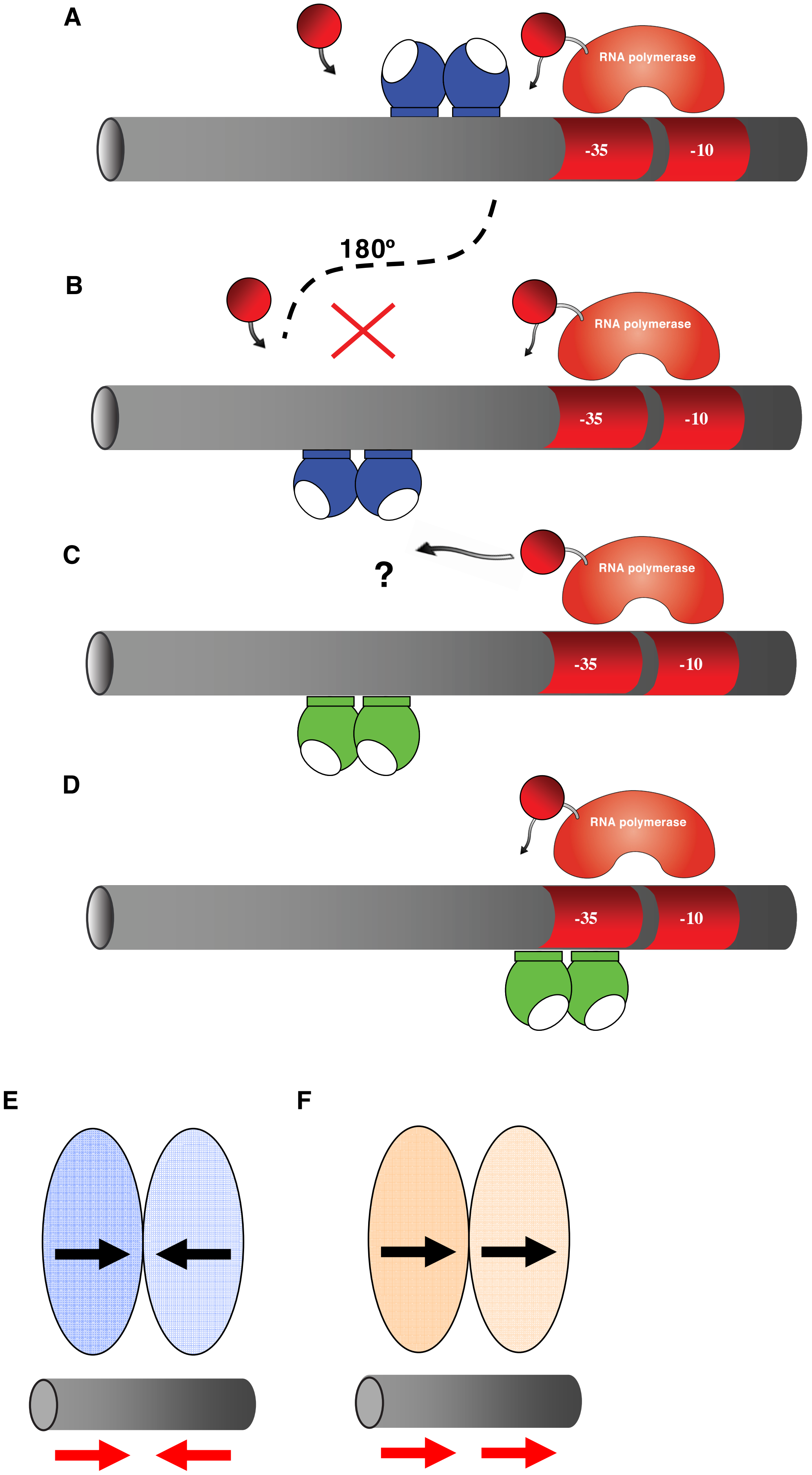

A
RNA polymerase
-35
-10
180º
B
RNA polymerase
-35
-10
C
?
RNA polymerase
-35
-10
D
RNA polymerase
-35
-10
E
F

Figure S5

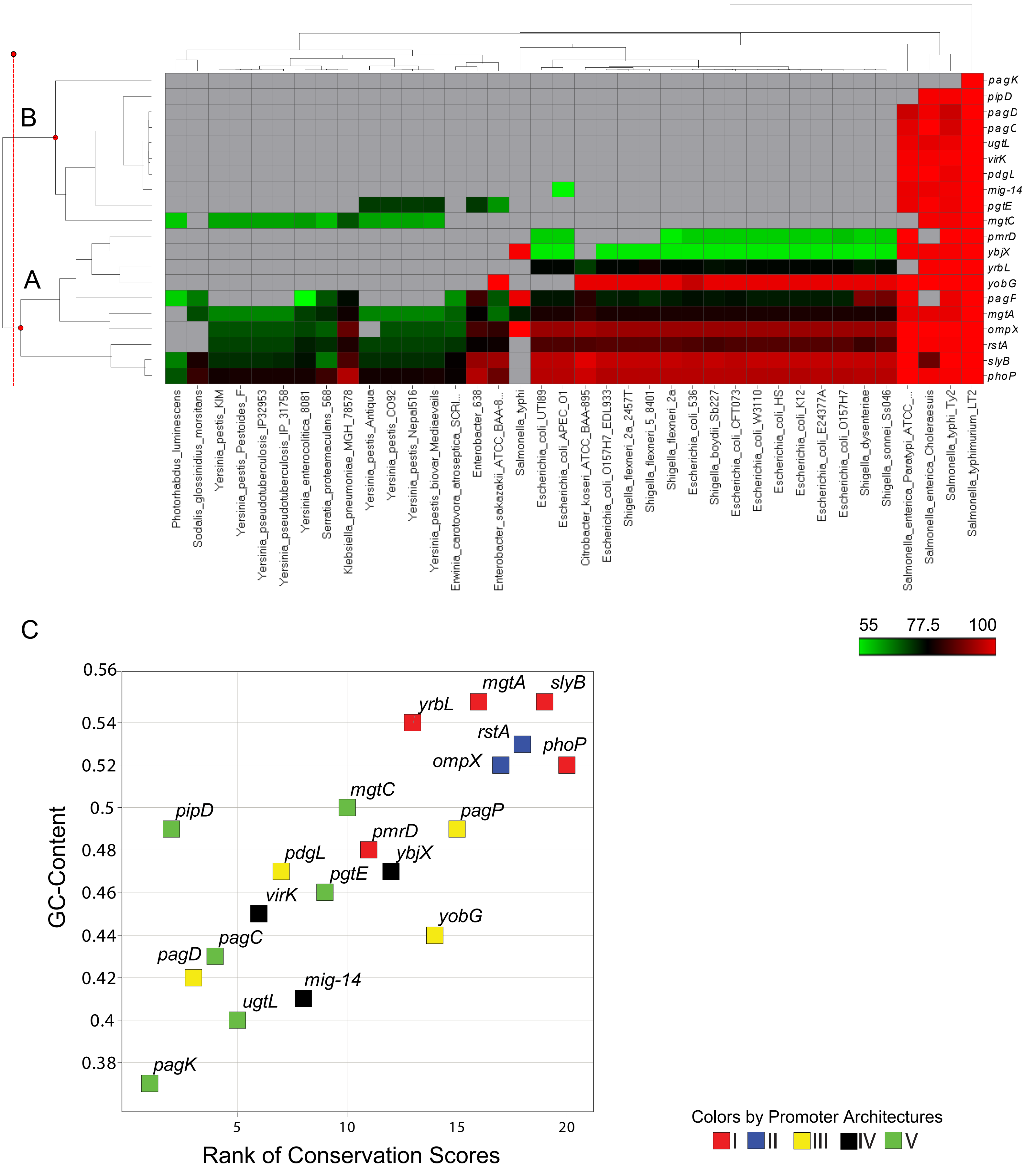

B
A
pagK
pipD
pagD
pagC
ugtL
virK
pdgL
mig-14
pgtE
mgtC
pmrD
ybjX
yrbL
yobG
pagF
mgtA
ompX
rstA
slyB
phoP
Photorhabdus_luminescens
Sodalis_glossinidius_morsitans
Yersinia_pestis_KIM
Yersinia_pestis_Pestoides_F
Yersinia_pseudotuberculosis_IP32953
Yersinia_pseudotuberculosis_IP_31758
Yersinia_enterocolitica_8081
Serratia_proteamaculans_568
Klebsiella_pneumoniae_MGH_78578
Yersinia_pestis_Antiqua
Yersinia_pestis_CO92
Yersinia_pestis_Nepal516
Yersinia_pestis_biovar_Mediaevails
Erwinia_carotovora_atroseptica_SCRI...
Enterobacter_638
Enterobacter_sakazakii_ATCC_BAA-8...
Salmonella_typhi
Escherichia_coli_UTI89
Escherichia_coli_APEC_O1
Citrobacter_koseri_ATCC_BAA-895
Escherichia_coli_O157H7_EDL933
Shigella_flexneri_2a_2457T
Shigella_flexneri_5_8401
Shigella_flexneri_2a
Escherichia_coli_536
Shigella_boydii_Sb227
Escherichia_coli_CFT073
Escherichia_coli_W3110
Escherichia_coli_HS
Escherichia_coli_K12
Escherichia_coli_E24377A
Escherichia_coli_O157H7
Shigella_dysenteriae
Shigella_sonnei_Ss046
Salmonella_enterica_Paratypi_ATCC_...
Salmonella_enterica_Choleraesuis
Salmonella_typhi_Ty2
Salmonella_typhimurium_LT2
55
77.5
100
C
GC-Content
0.56
0.54
0.52
0.5
0.48
0.46
0.44
0.42
0.4
0.38
5
10
15
20
Rank of Conservation Scores
mgtA
slyB
yrbL
rstA
phoP
ompX
mgtC
pipD
pagP
pmrD
pdgL
ybjX
pgtE
virK
yobG
pagC
pagD
mig-14
ugtL
pagK
Colors by Promoter Architectures
I
II
III
IV
V

Figure S6