Supplementary information Appendix for "Occurrence, Evolution and Functions of DNA Phosphorothioate Epigenetics in Bacteria"

Table S1. Analysis of consensus sequences in *P. fluorescens* pf0-1

		Downstream			
	5'-GGCC-3'	**A**	**C**	**G**	**T**
Upstream	**A**	4.4% (242/5,519)	2.2% (92/4,180)	4.8% (437/9,026)	2.7% (123/4,534)
	C	4.9% (659/13,344)	3.9% (248/6,391)	5.4% (554/10,234)	6.2% (557/9,026)
	G	3.8% (184/4,888)	1.5% (24/1,628)	5.1% (326/6,391)	4.5% (190/4,180)
	T	5.0% (478/9,488)	3.6% (174/4,888)	4.8% (643/13,344)	4.9% (271/5,519)

Table S2. Distribution of PT in different genome regions in *P. fluorescens* pf0-1

Regions	Total PT	Region number	Region total length (nt)	PT density (PT per 10^3 nt) #	Genomic PT density (PT per 10^3 nt)	*p* value (region vs genome)
ORF	2549	5722	5780682	0.44 ± 0.73		0.23
ORF promotor	318	5722	1143839	0.278 ± 1.161		1.59×10^{-6}***
rRNA	8	18	24427	0.238 ± 0.314		5.84×10^{-11}***
tRNA	2	73	5690	0.37 ± 3.163	0.404	0.24
rRNA promotor	0	18	3600	0		NA
tRNA promotor	0	73	12273	0		NA
Noncoding	2544	2045	6430618	0.336 ± 0.574		0.014*

#: mean ± SD, *: $p < 0.05$; ***: $p < 0.001$; NA: not applicable

Table S3. Differentially transcribed genes in TT-5

Gene ID	Function	Promoter PT sites	Internal PT sites	COG	Fold change (log_2)[a]
Pfl01_0738	DNA PT modification protein DndB	0	1	not found	-4.82
Pfl01_0739	DNA PT modification protein DndC	0	0	E	-17.39
Pfl01_0740	DNA PT modification protein DndD	0	1	L	-17.73
Pfl01_0741	DNA PT modification protein DndE	0	0	not found	-15.93
Pfl01_0900	Putative membrane protein	0	0	M	1.04
Pfl01_1169	Putative phage tail assembly protein	0	0	not found	1.06
Pfl01_1797	Putative xanthine dehydrogenase	0	0	F	1.64
Pfl01_2488	Hypothetical protein	0	1	not found	1.02
Pfl01_3496	Phage protein	0	0	not found	1.04
Pfl01_5657	N-acetylmuramoyl-L-alanine amidase	0	0	M	1.85

[a]Difference in the expression of TT-5 compared with that of wild-type *P. fluorescens* pf0-1; the "Internal PT sites" column indicates the number of PT-modified G_{PS}GCC motifs occurring in each gene's internal region. The COG categories are identified by capital letters as follows: E, amino acid transport and metabolism; L, replication; M, cell wall/membrane/envelope biogenesis; and F, nucleotide transport and metabolism.

Table S4. DNA templates for *in vitro* transcription

Gene ID	Sequence
Pfl01_1326	gtcaatttgccagcatgcgagtgcacggatagaaacccagtcagag*gccgctcATGGAATTCTTCGAAAA GTTAACCAGTCTTGCCGCCAAAGTTCGGTTGCAGAGCGCTGCAATCCAGACAGAA GAGGCGACAAAAAACGCGTTTGTCATGCCCTTCATCAGTACTGTTCTGGGTTACG ACGTATTCGACCCGACAGAGGTGACGCCGGAATTTGTTTGCGATATCGGCACGAA GAAGGGCGAAAAAATCGATTACGCCATCATGAAAGAGGGCGAAGTGCAGATCCTC ATCGAGTGCAAGAAAATCGGCGAGCCACTGCACATCAATCACGCCTCGCAG
Pfl01_1884	ttcgactttgccgccagcggcggtttctctacaatccgccgcagtttagccgaccccatccttagcggtcg*gcctggatcc ttttttcATGGACGATATTCAACAGCGCTGGCTGTGCGCCCTCTCTGCGCCAATGGCC GCGCTTAATACCGGTGCCGGGTATGACGACCCTGCCTTCTGCGACGACCGCTAT ATCGATCTGAAAGGCAGCTGGGGAATCGATGACCGGCGACAATTGTTCGACATG CTTCAATGGATGACGGACGACGGACATGCAAAGCACCTGAGTGGCGCCTATTCG GCGTGGCAACGTTGTCTGCCGAATGAATGGCAGCGCCTGCTCGAAGAGCTGAGC CCTCGCGAGCGTGTTCTGCATGAGTTTGCCAGTCGCAC
Pfl01_5242	tgttgattgaacgttcaatcaaaacaaaatagactggccctcaatccgatcggcgtttgtcgtccgtcggaag*gcctaag gagatctgcaagATGCCCAAGGTCGGTATGCAACCCATCCGCCGCCAACAACTGATCG AAGCCACTTTGCAGGCGGTTGATCAGGTCGGAATGGGGGACGCCAGCATTGCGC TGATCGCCCGTTTGGCCGGTGTCTCGAATGGCATCATCAGTCACTATTTTCAGGA CAAGAACGGCCTGATTGCCGCCACGATGCGGTATCTGATGAGCGTCCTCAGCGA GAACGTCACCGCGCGCCGTCAGGCGCTGGCAGACAGCAGCCCCCGGGCGCATC TGCAGGTGATCATCGAAGGCAAC
Pfl01_0914	ccgccaccgattacagggttgtggcataacccgatggcagcgcg*gcccggcactggacgaatggcaaagacgggg ctaaaacgaggcatgccccggagatttagATGAGTCCGCGCCAACGTTTCTTCGATTGTCTGC ACCGTTCACCGCCCGCGCTGTTCGAAGCGGCGTTGTGGATGGCCGCCGAGCAC GACAAACAAGCCGATCCCGAAGCGCTGTTGCAGGAATTCAAGGAACTGCAGCAG CGGGTCAGCTATGGCTTGCCGTTGCTGCCGGTCAGCGAGCTGGCGCAACCGCT GCTGCGGCGCATGACCGATCTGGGGTTTGCCCAGGATGATTTCGTACCGTTGC
Pfl01_3453	tcagggcggggccttcgttgcg*gccggcccgttcggcgattgaccggcgcgattcacgtctccaataaaaccaaggg aatgagtATGCAAGCCGAATTGATCGTCGACGCCCGCAACGCGGTGGGTGAAAGCC CGGTGTGGGTGCCGCAGGAAAATGCGCTCTACTGGGTCGATATCCCCAACGGTG GCCTGCAACGCTGGGATGCCGACACTGGCCATGTTCACGCCTGGAAGGCTCCCG AGATGCTCGCCTGCCTCGCCCGACACAGCCGGGGCGGCTGGGTGGCCGGTATG GAGAGCGGGTTCTTTCAACTGCGCCCCCACAGCGATGGCAGCCTTGA
Pfl01_5077	ccagttagtctgatcgcatacgacaaggtgttgatcaccgtgtcg*gccgtgaagaaattcgaggagctgctgggATG AACCAGGAACGCGTATTTAAAGTTCTGCTTGGCCCGCACGTTTCCGAGAAGGCTA CGGTTCTGGCAGACAAGAAAGGCCAGTTCGTTTTCAAGGTTGCTACTGACGCAAC CAAGCTGGAAATCAAGAAGGCCGTCGAAAGCCTGTTCAGCGTGAAAGTTGAGCG TGTTACTACCCTGAACGTTCTGGGTAAGAGTAAGCGCACCGCTCGCGGTCTGGG CAAGCGTAATGACTGGAAGAAGGCAGTTATCTCCCTTCAGCCAG
Pfl01_1383	gacgattcgcggatcatgggcctggggctgaagacgttgacggtgcaatgacgcagcgtttcgatggacgccggcccc gcggcctttctacactgcccgcaatctggggatcgggggcaatggATGAACCGCAATGAACTACGCAA GGATCCG*GCCGACATCAACCTGATGGTGGTGTTCGAGACCTTGATGCTCGAACG CAACGT
Pfl01_5297	gggtcagcgaatgcttgaagcaagcgatcagcgatattgcggcgactgatttgagcaatcgtctcggcgcgaaaaatc gcgaagcctccagccgcgtgcgtgaattgctgagggcgcaGTGGTGAACACGCCTCTGGATCCG*G CCCACGACCTGCTGTGGGGCATGACGCCGGCGCAGTTGCCGGCGGATGCGCC
Pfl01_0510	acatcaccgtccaagccttcggcaaaaccgcgcagtgcatcgagaaagccctgcacagttttagccagcgcggccaa tgacagctacaatccctcccctttgtgctctggagacctcacATGTCTGATTTCTCTGGCCTGGATCCG* GCCCTGGTGATCGAGCCGAGCGATCTGCTGCCGCGCCTCGAATCCCGCGATCT
Pfl01_4820	cctgccgattcctacaacccgtaaccgcactatctgacacttttgtcacctgcaccccgcgtgtttggcgctgtcacttgact aggctataagtcatgaactgtctgcatccgggattcccATGAATGACATCGCTCTGAGCGGTCTGGA TCCG*GCCAAGCAATTGGTACAGGCCGAACTGCTCACGGAAAAAAGCGCCCAGC ACGC

* Denotes chemically synthesized PT modification; the sequences in red are commercially synthesized; the sequences in blue are amplified by PCR; and the sequences in capitalized letters are predicted ORFs.

Table S5. Differential metabolites between *P. fluorescens* pf0-1 and TT-5

	Name	VIP	m/z	TT-5/WT fold change
ESI (+)				
H	Proline	1.52	116.0016	5.26
H	Phenylalanine	1.76	166.085	2.27
Na	C16 sphinganine	1.49	296.2567	2.04
Na	1-(9Z-nonadecenoyl)-glycero-3-phosphate (PA(19:1(9Z)/0:0))	1.86	473.2597	1.75
H	Oleic Acid	1.77	283.2658	1.64
H	1-Tridecanoyl-sn-glycero-3-phosphocholine (PC(13:0/0:0))	1.84	454.2918	1.61
H	Glyceric acid	1.38	107.0307	1.61
H	D-Glucose 6-phosphate	1.37	261.0386	1.61
Na	Phytosphingosine	1.81	340.2921	1.54
H	1-Undecanoyl-sn-glycero-3-phosphocholine (PC(11:0/0:0))	1.83	427.2657	1.52
H	Cysteine	1.8	144.0074	1.52
Na	1-Dodecanoyl-glycero-3-phosphoserine (PS(12:0/0:0))	1.87	464.2088	1.45
H	1-(1Z-hexadecenyl)-sn-glycero-3-phosphoethanolamine (PE(P-16:0/0:0))	1.77	438.2956	1.41
Na	Jasmonic acid	1.51	233.1121	1.28
H	1-(9Z-octadecenoyl)-sn-glycero-3-phospho-(1'-myo-inositol) (PI(18:1(9Z)/0:0))	1.53	599.3368	1.27
H	1-(11Z-eicosenoyl)-glycero-3-phospho-(1'-sn-glycerol) (PG(20:1(11Z)/0:0))	1.26	539.3497	1.25
H	1-Dodecanoyl-2-tridecanoyl-sn-glycero-3-phosphate (PA(12:0/13:0))	1.32	551.3574	1.25
Na	1-Hexadecanoyl-sn-glycero-3-phospho-(1'-sn-glycerol) (PG(16:0/0:0))	1.32	507.2732	1.23
H	5-HETE	1.35	321.0889	1.22
H	Kynurenine	1.32	209.0684	1.19
H	Sphinganine	1.6	302.3046	1.16
Na	1-(9Z-tetradecenoyl)-glycero-3-phosphoserine (PS(14:1(9Z)/0:0))	1.65	490.222	0.88
Na	Indole	1.79	140.0508	0.88
H	Cyclic GMP	1.64	346.1362	0.84
Na	1-Tetradecanoyl-glycero-3-phosphoserine (PS(14:0/0:0))	1.7	492.2434	0.83
H	1-Nonanoyl-sn-glycero-3-phosphocholine (PC(9:0/0:0))	1.35	398.2399	0.83
Na	N-(2-hydroxy-eicosanoyl)-tetradecasphing-4-enine (Cer(d14:1(4E)/20:0(2OH)))	1.92	576.508	0.83
Na	1-Propionyl-sn-glycero-3-phosphocholine (PC(3:0/0:0))	1.31	336.1207	0.82
H	Cytosine	1.37	112.0442	0.74
Na	MG(0:0/18:2(9Z,12Z)/0:0)	1.3	377.2615	0.72
H	C17 sphingosine-1-phosphate	1.83	366.2335	0.67
H	Taurine	1.74	125.9449	0.62
Na	Glycine	1.43	98.0219	0.50

H	Citric acid	1.56	215.0132	0.45
Na	1,2-Dihexanoyl-sn-glycero-3-phosphoethanolamine (PE(6:0/6:0))	1.4	434.1877	0.33
Na	1-Octadecanoyl-sn-glycero-3-phosphate (PA(18:0/0:0))	1.44	461.2582	0.20
Na	Quinic acid	1.89	215.0559	0.14
Na	N-(heptadecanoyl)-sphinganine (Cer(d18:0/17:0))	1.38	576.5245	0.06
ESI (-)				
H	Leukotriene A4	2.17	492.2273	5.88
H	Docosanoic acid	1.92	659.4733	5.26
H	3-Dehydrogulonate	1.67	278.4681	2.22
H	2-Aminoadipate semialdehyde	1.67	186.2861	2.22
H	Phosphoribosylamine	1.55	297.4832	2.08
H	Uridine triphosphate（UTP）	1.92	855.5725	1.18
H	17beta-Estradiol 3-(beta-D-glucuronide)	2.06	773.4253	1.16
H	3a-Hydroxy-5b-pregnane-20-one	1.55	492.2789	1.09
H	L-methionine	1.49	189.4449	0.50
H	Progesterone	1.57	478.9468	0.47

Table S6. Quantification of genes related to proline metabolism in *P. fluorescens* pf0-1 and TT-5 by real-time RT-PCR

	Genes	pf0-1	TT-5
Average Ct ± SD	*Ornithine cyclodeaminase*	23.33 ± 1.27	22.89 ± 1.20
	proA	23.68 ± 1.61	23.99 ± 2.05
	proB	19.52 ± 1.52	19.27 ± 1.30
	proC	18.76 ± 1.60	18.33 ± 1.13
	putA	17.68 ± 1.64	17.47 ± 0.93
	rpoD	16.44 ± 1.60	16.61 ± 1.26
ΔCt ± SD	*Ornithine cyclodeaminase*	6.89 ± 0.33	6.28 ± 0.064
	proA	7.24 ± 0.0007	7.38 ± 0.78
	proB	3.08 ± 0.088	2.66 ± 0.036
	proC	2.32 ± 0.0002	1.72 ± 0.14
	putA	1.25 ± 0.038	0.86 ± 0.34
Fold change relative to *ropD* $2^{-\Delta\Delta Ct}$ (TT-5 vs pf0-1)	*Ornithine cyclodeaminase*	1.54 ± 0.29	
	proA	1.07 ± 0.65	
	proB	1.34 ± 0.11	
	proC	1.52 ± 0.15	
	putA	1.33 ± 0.34	

Table S7. Bacterial strains and plasmids used in this study

Strains or plasmids	Genotype or description	Sources or reference
E. coli		
WM3064	*thrB1004 pro thi rpsL hsdS lacZΔM15RP4–1360Δ(araBAD)567ΔdapA1341::[erm pir(wt)]*	(1)
Trans1-T1	F$^-$ φ80(*lacZ*)ΔM15*Δlac*X74h*sd*R(r_k^-, m_k^+)*ΔrecA*1398*end* A1 *ton*A	TransGene Biotech
P. fluorescens		
pf0-1	Wild type, Amp^r	(2)
TT-5	pf0-1 derivative, *dndBCDE* deletion mutant	This work
Plasmid		
pEASY^{TM} Blunt Zero	Cloning vector, Amp^r, Kan^r	TransGene Biotech
pSR47s	*sacB*-containing suicide vector (containing the R6K replication origin), kan^r	(3)
pWHU3345	pEASY-blunt zero derivative with a 1,725-bp NotI fragment carrying an in-frame deletion of *dndBCDE*	This work
pWHU3350	Suicide plasmid pSR47s derivative with a 1,725-bp NotI fragment carrying an in-frame deletion of *dndBCDE*	This work
pWHU3636	pDSK519 derivative with a 5.3-kb NotI-BamHI fragment carrying *dndBCDE* of pf0-1	This work

Table S8. PCR and DNA sequencing primers used in this study

Primer	5'-3'
Construction of TT-5 with in-frame deletion of *dndBCDE*	
spf01	GCGGCCGCTTAGGATGTGGCAGGTGACT
spf02	TGAAAGCCCTCTGTGGACATGAGGCATAAGCTCTTCGTTGT
spf03	ACAACGAAGAGCTTATGCCTCATGTCCACAGAGGGCTTTCA
spf04	GCGGCCGCGCCTTAATCTTCGCTTCAGCAT
spfup	GATGTCTTCGCTTCCG
spfdown	GAAGTAGAGCCCGAAACA
***In vitro* transcription**	
Pfl01_1326-qPCR-F	ATCCAGACAGAAGAGGCGAC
Pfl01_1884-qPCR-F	CAATGGCCGCGCTTAATACC
Pfl01_5242-qPCR-F	TGATCGAAGCCACTTTGCAG
Pfl01_0914-qPCR-F	GATTGTCTGCACCGTTCACC
Pfl01_3453-qPCR-F	GGAAAATGCGCTCTACTGGG
Pfl01_5077-qPCR-F	AGAAAGGCCAGTTCGTTTTCA
Pfl01_1383-qPCR-F	TGGATGAACCGCAATGAACTAC
Pfl01_5297-qPCR-F	GCAGTGGTGAACACGCCTCT
Pfl01_0510-qPCR-F	ATGTCTGATTTCTCTGGCCTGG
Pfl01_4820-qPCR-F	CCATGAATGACATCGCTCTG
qPCR-R (Tag)	AATTGGTGACACTCAGGCAC
Pfl01_1326-reverse transcription	AATTGGTGACACTCAGGCACCAAACAAATTCCGGCGTCAC
Pfl01_1884-reverse transcription	AATTGGTGACACTCAGGCACGCATGTCGAACAATTGTCGC
Pfl01_5242-reverse transcription	AATTGGTGACACTCAGGCACCAATCAGGCCGTTCTTGTCC
Pfl01_0914-reverse transcription	AATTGGTGACACTCAGGCACTGCTGCAGTTCCTTGAATTCC
Pfl01_3453-reverse transcription	AATTGGTGACACTCAGGCACGAGCATCTCGGGAGCCTTC
Pfl01_5077-reverse transcription	AATTGGTGACACTCAGGCACGGGTAGTAACACGCTCAACT
Pfl01_1383-reverse transcription	AATTGGTGACACTCAGGCACTCTCGAACACCACCATCAGG
Pfl01_5297-reverse transcription	AATTGGTGACACTCAGGCACAACTGCGCCGGCGTCATG
Pfl01_0510-reverse transcription	AATTGGTGACACTCAGGCACAGCAGATCGCTCGGCTCGAT
Pfl01_4820-reverse transcription	AATTGGTGACACTCAGGCACGTGAGCAGTTCGGCCTGTA
Construction of the pWHU3636 plasmid	
spf-B-up	CTAGTATGACGTCTGTCGCACCTGCTTGATCGCGGCCGCAAT GGCACGGTAGGTAGC
Spf-E-down	CCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCT CACAAAAGGTGGCGA
Proline metabolism	
proB-F	CTTCTGACTCACGACGACCT
proB-R	CGGATTTCGTCAGTGACCAC
proA-F	TGCAACACCTTGAACGTCTG
proA-R	GCCTTCGACGATGTGCAATT
putA-F	ACAAAGGTCGCAACGTGTAC
putA-R	AAGATCTCGCGTTGCAGTTC
proC-F	GTCTGGAAGCTGCACACATC
proC-R	CTTGACCGCCAGTACCACTA
Ornithine cyclodeaminase-F	AAAGCCTCAACGACATGCAG
Ornithine cyclodeaminase-R	CTTTTCCGCTGAGCAGATCC

p=0.003

DndBCD

DndFGH

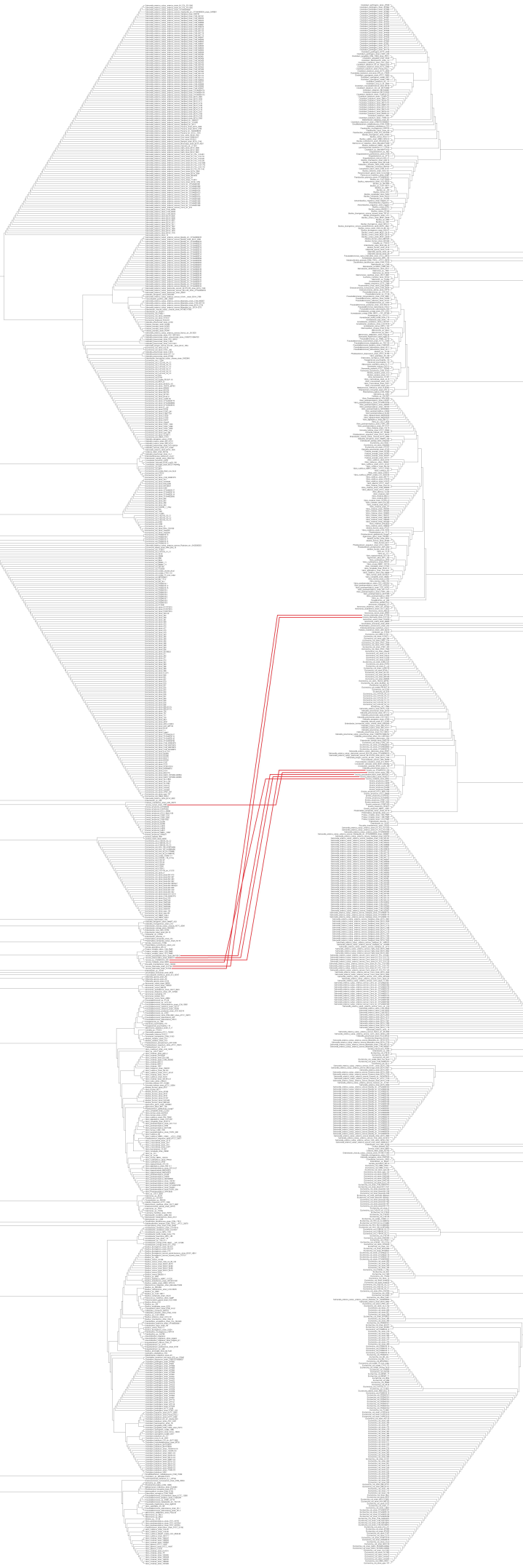

Figure S1. Comparison of the phylogenies of the M and R components of PT systems. Tanglegram for DndBCD proteins (left) and DndFGH (right) proteins based on 734 strains with complete PT R-M systems. The tanglegram was prepared using Dendroscope 4.4.4. The topological difference between the DndBCD and DndFGH trees is significantly small (p=0.003) compared with that between tree pairs formed by randomly shuffled strains. Examples of phylogenetically divergent DndBCD and DndFGH from one strain are connected by red lines.

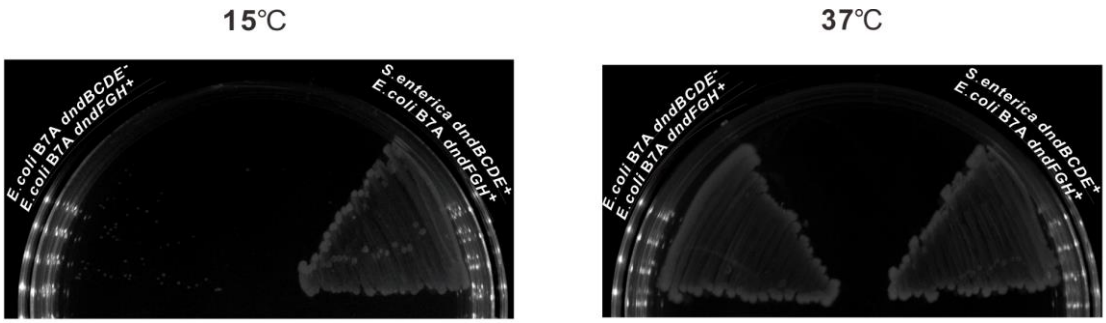

Figure S2. The PT M genes from *S. enterica* serovar Cerro 87 rescued the DndFGH-induced growth inhibition in a PT-deficient *E. coli* B7A mutant. Effects of pJTU1238, expressing DndBCDE from *S. enterica* serovar Cerro 87, on the growth of *E. coli* B7A Δ*dndBCDE* at 37 and 15°C on LB-agar plates. DndFGH-mediated growth inhibition is temperature dependent (4). *E. coli* B7A Δ*dndBCDE* harboring empty pBluescript SK+ plasmid was used as the control.

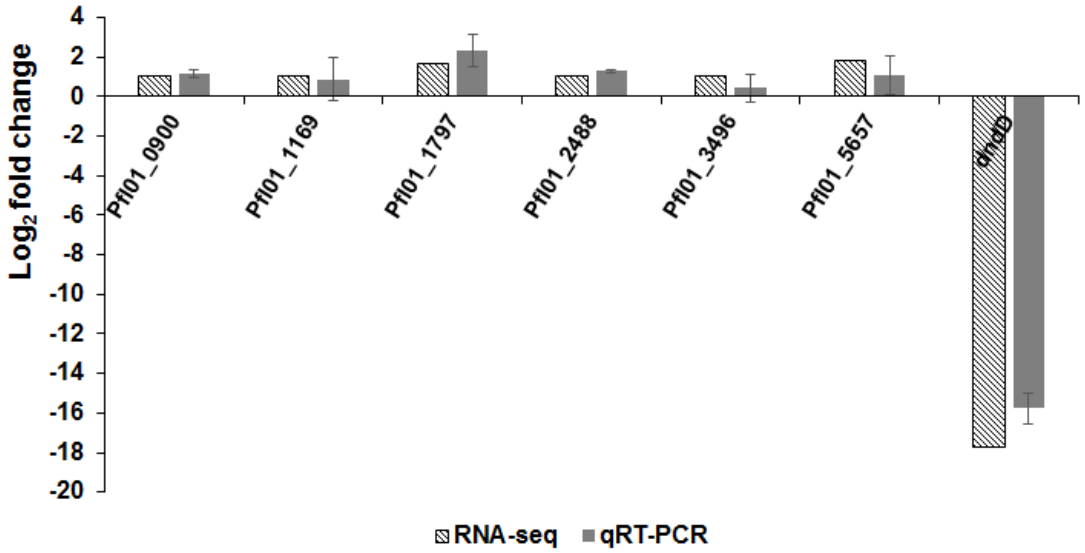

Figure S3. qRT-PCR validation of differentially transcribed genes. Analysis of the seven differentially transcribed genes as determined by RNA-Seq (patterned bars) or qRT-PCR (solid bars). The levels are presented relative to the values obtained for the wild-type pf0-1 strain. The values represent the average gene expression values ± SDs from three independent qRT-PCR experiments. The housekeeping gene, *rpoD*, which encodes sigma factor σ^{70} was used as the reference for RT-PCR.

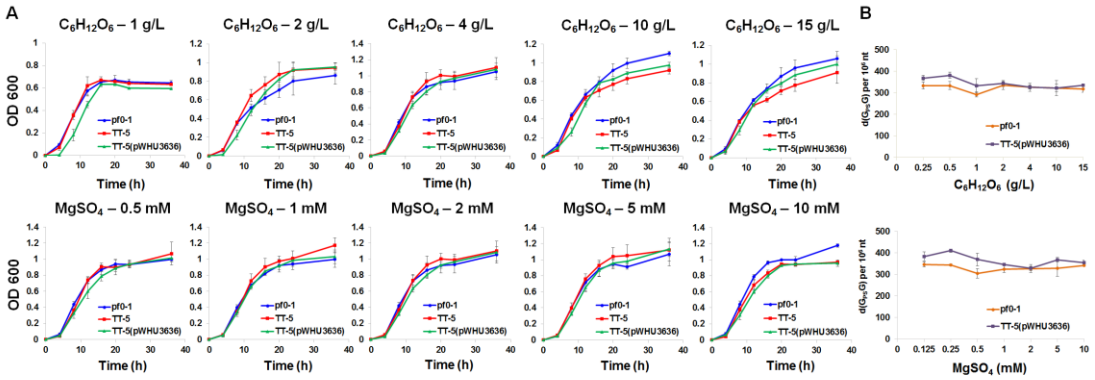

Figure S4. (A) Growth of wild-type *P. fluorescens* pf0-1 and the mutant TT-5 and TT-5(pWHU3636) cells cultured in M9 minimal medium supplemented with different concentrations of glucose and sulfate as the carbon and sulfur sources, respectively. TT-5 displayed similar growth profiles to those of the wild-type pf0-1 and TT-5(pWHU3636). pWHU3636 is the derivative of the pDSK519 plasmid, which possesses *dndBCDE* from pf0-1. (B) Different concentrations of carbon and sulfur sources had no apparent impact on the frequency of PT-modified d(G_{PS}G) in pf0-1 and TT-5(pWHU3636).

REFERENCES

1. Dehio C & Meyer M (1997) Maintenance of broad-host-range incompatibility group P and group Q plasmids and transposition of Tn5 in Bartonella henselae following conjugal plasmid transfer from *Escherichia coli*. *J Bacteriol* 179(2):538-540.
2. Yao F, Xu T, Zhou X, Deng Z, & You D (2009) Functional analysis of spfD gene involved in DNA phosphorothioation in *Pseudomonas fluorescens* Pf0-1. *FEBS Lett* 583(4):729-733.
3. Merriam JJ, Mathur R, Maxfield-Boumil R, & Isberg RR (1997) Analysis of the *Legionella pneumophila* fliI gene: intracellular growth of a defined mutant defective for flagellum biosynthesis. *Infect Immun* 65(6):2497-2501.
4. Chen C, *et al.* (2017) Convergence of DNA methylation and phosphorothioation epigenetics in bacterial genomes. *Proc Natl Acad Sci USA* 114(17):4501-4506.