2686–2697 *Nucleic Acids Research, 2006, Vol. 34, No. 9*
doi:10.1093/nar/gkl365

Genome prediction of PhoB regulated promoters in *Sinorhizobium meliloti* and twelve proteobacteria

Ze-Chun Yuan, Rahat Zaheer, Richard Morton and Turlough M. Finan*

Center for Environmental Genomics, Department of Biology, McMaster University, Hamilton, Ontario, Canada, L8S 4K1

Received March 2, 2006; Revised March 29, 2006; Accepted April 26, 2006

ABSTRACT

In proteobacteria, genes whose expression is modulated in response to the external concentration of inorganic phosphate are often regulated by the PhoB protein which binds to a conserved motif (Pho box) within their promoter regions. Using a position weight matrix algorithm derived from known Pho box sequences, we identified 96 putative Pho regulon members whose promoter regions contained one or more Pho boxs in the *Sinorhizobium meliloti* genome. Expression of these genes was examined through assays of reporter gene fusions and through comparison with published microarray data. Of 96 genes, 31 were induced and 3 were repressed by Pi starvation in a PhoB dependent manner. Novel Pho regulon members included several genes of unknown function. Comparative analysis across 12 proteobacterial genomes revealed highly conserved Pho regulon members including genes involved in Pi metabolism (*pstS*, *phnC* and *ppdK*). Genes with no obvious association with Pi metabolism were predicted to be Pho regulon members in *S.meliloti* and multiple organisms. These included *smc01605* and *smc04317* which are annotated as substrate binding proteins of iron transporters and *katA* encoding catalase. This data suggests that the Pho regulon overlaps and interacts with several other control circuits, such as the oxidative stress response and iron homeostasis.

INTRODUCTION

Dissection of regulatory networks that control gene transcription is among the primary goals of the post-genomic era of biology. Whether gene expression is measured from microarrays or reporter gene fusions or other methodologies, it is generally not possible to distinguish between the direct and indirect modulation of transcription. Bioinformatic approaches to identify the regulatory networks have included the design of algorithms for genome-wide prediction of conserved regulatory DNA binding motifs (1,2). A promising approach in the delineation of transcriptional networks lies in combining genomic scanning or *in silico* analysis with experimental transcription data obtained from cells grown under diverse experimental conditions (1,3–12). In this report, we combine *in silico* prediction with experimental data obtained from reporter gene fusions and through comparisons with published microarray data. We also explore cross-species comparative genomics as a tool to identify genes whose expression is controlled by a transcriptional regulator, PhoB, in response to the phosphate starvation.

Inorganic phosphate (Pi) plays key roles in cells. In ATP, it is involved in energy metabolism, in protein phosphorylation it is responsible for regulation of transcription and many other cellular processes including chemotaxis and cell division, and perhaps most importantly, Pi is a major structural component of nucleic acids and membrane phospholipids. In many gram-negative bacteria, the transport and metabolism of Pi and phosphorous containing compounds is regulated at the transcriptional level by a two-component PhoR-PhoB signal transduction system. The Pho regulon consists of genes or operons regulated by PhoB and this has been well studied in *Escherichia coli* (13–15). Under Pi limiting conditions, the PhoR histidine kinase sensor undergoes autophosphorylation and subsequently donates its phosphate group to its cognate response regulator PhoB. Phosphorylated PhoB (PhoB-Pi) then modulates transcription of its targets by binding to a highly conserved 18 nt DNA sequence called the Pho box (or PhoB binding motif) which usually overlaps the −35 region of PhoB-regulated promoters (16,17). The majority of identified Pho boxes essentially comprised two 7 nt direct repeats of 5′-CTGTCAT-3′ separated by a conserved 4 nt spacer in the middle. It was postulated that the PhoB and Pho box binding complex interacts with the σ^{70} subunit of RNA polymerase to control transcription initiation (18–23). Over the past 30 years, about 30 Pho regulon members, which predominantly encompass an ensemble of genes

*To whom correspondence should be addressed. Tel: +905 525 9140, ext. 22932; Fax: +905 522 6066; Email: finan@mcmaster.ca

© The Author 2006. Published by Oxford University Press. All rights reserved.

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please contact journals.permissions@oxfordjournals.org

involved in Pi uptake and metabolism, have been identified in *E.coli* as reviewed by Wanner (14).

We are studying the gram negative α-proteobacterium *Sinorhizobium meliloti*. This organism forms N_2-fixing root-nodules on alfalfa (*Medicago sativa*) and its genome is unusual as in addition to a 3.4 Mb chromosome it contains two megaplasmids 1.3 and 1.7 Mb in size. Previous studies have identified several Pho regulon members in *S.meliloti* including the *pstSCAB* and *phoCDET* operons which encode ABC-type high affinity transport systems for Pi and in the case of *phoCDET* likely phosphonates (24–27). The *orfA-pit* operon encodes a low affinity Pi transport system whose expression is negatively regulated by PhoB (28). Other Pho regulons include the *exp*, *phn* and *pta-ackA* operons (16,29,30). Using DNA microarray and promoter analysis, Krol and Becker (31) identified several novel putative Pho regulon members including *afuA* which is annotated as an iron transport binding protein. In ongoing studies to understand the response of *S.meliloti* to Pi limitation, we constructed a Pho box weight matrix based on known *E.coli* and *S.meliloti* PhoB binding sites and used this matrix to predict new PhoB binding sites in the *S.meliloti* genome. Expression of predicted Pho regulon members then was examined through the analysis of transcriptional reporter gene fusions and through the previously reported microarray data (31). The frequency weight matrix was also employed to predict PhoB binding motifs across 12 closely related proteobacterial genomes with a goal to identifying a common set of PhoB regulated genes as might be expected from a conserved biological response to Pi-limitation.

MATERIALS AND METHODS

Construction of the Pho box weight matrix for prediction of PhoB binding sites

A total of fifteen known Pho boxes from *S.meliloti* and *E.coli* were used for weight matrix construction. Five of those Pho box sequences were collected from previously identified PhoB binding sites from *S.meliloti*. Of those five, four were from *S.meliloti* strain 1021 including one PhoB binding site upstream *orfA-pit;* two sites from the *phoC* promoter; one from the *phnG* promoter (24,28,32) and one Pho box was taken from the *orfA-pta-ackA* promoter of *S.meliloti* strain 104A14 (16). Ten PhoB binding sites from *E.coli* were *phoA, phoB, phoE, phoH, phnC, pstS1, pstS2, ugpB1, ugpB2* and *ugpB3* (18,33–35) (see Table 1). Following their alignment a matrix was constructed from the relative frequencies of A, T, C or G at each position of the 18 nt Pho box sequence (Table 1). This matrix was used to determine an information-based measure of potential binding sites according to the method of Schneider *et al*. (36). An 18 bp window was moved over the entire genome on both strands and the score (S_i) at each nucleotide position (having base i) was calculated according to $S_i = (1/18)\ \Sigma_j\ [2 + \log2(F_{ij})]$, where F_{ij} is the frequency matrix for base i at position j. This score, which ranges from −2.62 (the score of the worse match) to 1.39 (the score of the consensus sequence), is a measure of the information content of a potential binding site measured against the example set. The lowest example score, that of *orfA-pit*, is 0.36 and a threshold of 0.35 was used to define a 'hit'. A scan of the entire *S.meliloti* genome produced about 1500 hits on each strand. These were filtered to retain only those that were between −500 to +100 bp on the coding strand from an annotated translational start site.

Generation of *gusA* transcriptional gene fusions to the PCR amplified Pho box containing promoters

To construct the *gusA* reporter gene fusions to the Pho box containing promoters, each promoter region was PCR amplified using the primers as listed in Supplementary Table 2, and the PCR amplified promoter fragments were digested with appropriate restriction enzymes and cloned into either pFUS1 vector which is a broad host replicable vector containing promoterless *gusA* (*uidA*) gene (37) or into a suicide

Table 1. Pho-box matrix

Position	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18
S.meliloti orfA-pta:	T	T	G	T	C	A	A	A	C	C	G	C	C	G	T	A	A	C
S.meliloti phnG:	A	T	G	T	C	A	C	A	A	G	C	C	T	G	T	C	A	T
S.meliloti phoC1:	C	T	G	A	C	A	C	T	G	C	G	C	T	T	T	C	A	T
S.meliloti phoC2:	C	T	G	T	T	A	C	A	G	A	A	C	C	T	A	C	A	C
S.meliloti orfA-pit:	C	T	G	T	G	G	G	A	A	A	G	C	C	G	T	T	T	T
E.coli phoA	C	T	G	T	C	A	T	C	A	C	T	C	T	G	T	C	A	T
E.coli phoB	C	T	G	T	C	A	T	A	A	A	G	T	T	G	T	C	A	C
E.coli phoE	C	T	G	T	A	A	T	A	T	A	T	C	T	T	T	A	A	C
E.coli phoH	C	T	G	T	C	A	T	C	A	C	T	C	T	G	T	C	A	T
E.coli phnC	C	T	G	T	T	A	G	T	C	A	C	T	T	T	T	A	A	T
E.coli pstS1	C	T	G	T	C	A	T	A	A	A	A	C	T	G	T	C	A	T
E.coli pstS2	C	T	T	A	C	A	T	A	T	A	A	C	T	G	T	C	A	C
E.coli ugpB1	T	T	G	T	C	A	T	C	T	T	T	C	T	G	A	C	A	C
E.coli ugpB2	C	T	A	T	C	T	T	A	C	A	A	A	T	G	T	A	A	C
E.coli ugpB3	A	A	G	T	T	A	T	T	T	T	T	C	T	G	T	A	A	T
Frequencies adjusted by adding 0.03 for zero count																		
A	0.13	0.06	0.07	0.13	0.07	0.84	0.07	0.58	0.4	0.53	0.27	0.07	0.03	0.03	0.13	0.32	0.88	0.03
T	0.13	0.88	0.07	0.81	0.2	0.07	0.6	0.19	0.27	0.13	0.33	0.13	0.75	0.25	0.81	0.07	0.06	0.03
C	0.71	0.03	0.07	0.03	0.67	0.03	0.2	0.19	0.2	0.27	0.13	0.77	0.19	0.03	0.03	0.58	0.03	0.44
G	0.03	0.03	0.84	0.03	0.07	0.07	0.13	0.03	0.13	0.07	0.27	0.03	0.03	0.69	0.03	0.03	0.03	0.03
Consensus sequence	C	T	G	T	C	A	T	A	A	A	T	C	T	G	T	C	A	T

S.meliloti and *E.coli* PhoB binding sites were extracted from the following references: (16,18,24,28,32,33–35).

plasmid pTH1360 [modified pVO155 (38) by replacement of gusA coding and upstream sequences with the ones in pFUS1]. The corresponding gene fusion plasmids were verified by sequencing and subsequently introduced into *S.meliloti* wild-type strains RCR2011 and its derivative RmP559 (RCR2011, *PhoB*$_3$::*TnV*) strains or RmP110 and RmH852 (Rm1021, *phoB*$_3$*::Tn5-233*) by tri-parental mating using MT616 as the helper strain as described previously (39).

β-Glucuronidase and alkaline phosphatase assays

To determine the expression of the predicted Pho box containing genes or operons in response to Pi limitation, the *S.meliloti* wild-type strains and its *PhoB* mutant harbouring the plasmid borne promoter::*gusA* gene fusion were inoculated in 2 ml LBmc containing 2.5 μg/ml tetracycline and grown overnight aerobically at 30°C to OD_{600} of ∼1.0. Luria–Bertani (LB) broth was supplemented with 2.5 mM $MgSO_4$ and 2.5 mM $CaCl_2$ (LBmc) (40). A total of 0.5 m of cultures were spun down in a 1.5 ml microcentrifuge tube, washed twice in 1 ml phosphate free MOPS minimal medium (P0 medium) and resuspended in 250 μl of the P0 medium. MOPS-buffered minimal medium contains 40 mM morpholinopropane sulfonic acid/20 mM potassium hydroxide; 20 mM NH_4Cl; 2 mM $MgSO_4$; 2 mM $CaCl_2$; 100 mM NaCl; 15 mm filter-sterilized glucose as carbon source and supplied with 0.3 μg/ml biotin and 10 ng/ml $CoCl_2$ (24,41). Ten microliter aliquots of washed cells were subcultured into 5 ml of P0 medium, or MOPS minimal medium supplied with 2 mM KH_2PO_4 (P2 medium). After 32 h incubation at 30°C, 2 ml cultures were spun down at 10 000 r.p.m. for 1 min and resuspended in 1 M Tris–HCl (pH 8.0) for alkaline phosphatase assays as described by Bardin *et al*. (24), and 3 ml cultures were left for β-glucuronidase assays according to the protocol described by Reeve *et al*. (37).

RESULTS AND DISCUSSION

Weight matrix prediction of potential PhoB regulated genes

A weight matrix to identify potential PhoB binding sites was generated from five *S.meliloti* and ten *E.coli* PhoB box example sequences of 18 nt length (Table 1). The nucleotide frequency matrix was used to calculate an information-based score for potential binding sites in a scan of the *S.meliloti* genome. Putative PhoB binding sites were defined by a score of greater than 0.35 and a location between +100 and −500 nt of the translational start codon on the transcribed strand of an annotated gene (see Materials and Methods). One hundred and three putative PhoB binding sites were found and are shown with their downstream annotated genes in Supplementary Table S1. Seven of these promoter regions contained two putative PhoB boxes, so that 96 distinct genes were found. Three out of four genes whose Pho boxes were used for matrix construction were also among those 96 genes. No orthologue of *orfA-pta-ackA* from *S.meliloti* strain 104A14 was found in Rm1021 strain.

The threshold score (0.35) used to identify putative PhoB binding sites was derived from the lowest score (that of *orfA-pit*) among the example sequences. With this threshold, 18 of the top 20 scores were upstream of genes found to be induced by phosphate starvation, in a PhoB-dependent manner, by *gusA* fusion analysis in *S.meliloti* (see next section). However, most putative PhoB binding sites with scores above the cut-off level did not show phosphate-dependent regulation of transcription. Possible explanations in addition to false positives are that the matrix method did not include other important features of a PhoB binding site, such as appropriately positioned −10 and −35 promoter elements. It is also possible that some genes with PhoB binding sites require interaction with additional regulatory proteins before the gene can be regulated by phosphate limitation.

Blanco *et al*. (23) showed that the C-terminal domain of PhoB interacts with a 22 bp region of dsDNA that consists of two direct repeats of 11 bp. Each 11 bp repeat has a conserved 7 bp region (consensus, CTGTCAT) followed by a less conserved 4 bp segment. Our weight matrix is comprised of two conserved 7 bp repeats separated by a single, less conserved 4 bp spacer, and omits the terminal 4 bp segment. However, this terminal segment is not well conserved (23) and will therefore contribute little to the weight matrix score. Furthermore, our weight matrix will reliably identify overlapping PhoB sites provided that they are separated by 4 bp 'spacers' and individually have component scores greater than 0.35.

Experimental validation of the predicted Pho regulon members by analysis of transcriptional gene fusions

To directly examine whether the *S.meliloti* genes identified by the frequency matrix were subject to phosphate-dependent regulation, we generated transcriptional reporter gene fusions to seventy-two of these candidate genes and examined their expression in defined MOPS-buffered minimal medium during growth under Pi-excess (2 mM Pi) and Pi-starvation (no Pi added) conditions (see Materials and Methods). Gene expression in a wild-type *phoB*$^+$ background was compared with expression in an otherwise isogenic *phoB*$^-$ background (Table 2). Eighteen of the 72 promoter gene fusions were induced upon Pi-starvation in a PhoB-dependent fashion (Table 2). In addition, regardless of the media Pi concentration, gene fusions to *smb20427* (putative amino acid ABC transport system), *smc02886* and *smc02675* (*rrna*) showed 3-, 2- and 10-fold more expression respectively in the wild-type background relative to the *phoB*$^-$ background.

Three reporter gene fusions were found to be repressed upon Pi-starvation in a PhoB-dependent manner. These were *smc00801* (transmembrane protein of unknown function), *smc02601* (*nadABC*) and *smc02862* (*orfA-pit*). In the wild-type background these fusions were expressed at higher levels in media containing 2 mM Pi than in Pi-starved cells. Also, in the *phoB* mutant background the expression level was elevated and did not alter with media Pi. We have previously reported that expression of the low-affinity Pi-transport system encoded by the *orfA-pit* genes is repressed by PhoB (28). The repression of *smc02601- smc02602- smc02603* (*nadABC*) expression suggests that the observed down regulation of NAD^+ synthesis in *S.meliloti* possibly corresponds to a slight down regulation of *smc00161* expression that appears to occur upon Pi-limited growth (Table 2). The

Table 2. β-Glucuronidase activities[a] from predicted Pho-regulon *gusA* gene fusions[b]

Gene[c]	Wild-type 0 Pi	Wild-type 2 mM Pi	*phoB*$^-$ 0 Pi	*phoB*$^-$ 2 mM Pi	Distance	Score
SMa0041	87.4 ± 1.3	110.3 ± 5.2	70.7 ± 4.6	113.2 ± 4.5	−101, −90	0.36, 0.48
SMa0302	88 ± 4.6	123.6 ± 8.2	134.6 ± 7	187.7 ± 1.2	−21	0.45
SMa0567	139.7 ± 4.8	66 ± 2.6	116.7 ± 5.4	65.9 ± 18.5	−130	0.37
SMa0570	220.7 ± 3.5	155.9 ± 3.3	169.3 ± 33.5	128.4 ± 35.9	−41	0.38
SMa1355	275.2 ± 44.2	231.2 ± 40.2	229.9 ± 21.8	277.4 ± 24.3	−160	0.41
SMa1456	726.2 ± 14.2	763 ± 41.2	804.2 ± 56	858.5 ± 34.6	−30	0.45
SMa1836	139.1 ± 13.6	162.9 ± 9.4	122.1 ± 10.8	146.4 ± 23.8	−285	0.39
SMa2012	101.1 ± 6	100.1 ± 1.4	112.2 ± 14	102.8 ± 23.3	−381	0.39
SMa2025	132 ± 0.7	206.5 ± 21.6	121.8 ± 2.8	207.8 ± 13.6	−22	0.36
SMa2063	37.7 ± 19.5	31.4 ± 17.1	55.2 ± 12.8	75.1 ± 7.9	+349	0.55
SMb20106	592.2 ± 8.7	915 ± 52.5	460.6 ± 23.3	910.4 ± 84.1	−60	0.43
SMb20410	42.6 ± 24.7	48.1 ± 16.2	70.2 ± 13.3	87.5 ± 12.8	−6	0.44
SMb20427	1895.2 ± 78.2	1943.1 ± 78.7	671.8 ± 34.2	606 ± 43	−129	0.37
SMb20483 (*crp*)	95 ± 45.5	92.7 ± 24.5	92.1 ± 15.3	102.2 ± 12.7	−80, −62	0.36, 0.37
SMb20493	88 ± 14.8	131.7 ± 38.6	156.7 ± 16	216.4 ± 10.6	−41	0.63
SMb20759(*phnG*)*	1622.6 ± 25.9	136.6 ± 8.5	72.8 ± 4.1	115.4 ± 5.9	−66	0.40
SMb20824	1062.7 ± 54.8	1974.06 ± 229.8	996.45 ± 10.2	1709.57 ± 76.1	−154	0.38
SMb20843*	1112 ± 149	239.7 ± 11.1	179.5 ± 28.5	200.5 ± 21.8	−340	0.38
SMb20876*	2092.4 ± 134	44.5 ± 19.6	76.8 ± 13.7	69.9 ± 7.5	−98	0.43
SMb20935	722.8 ± 29.5	1182.3 ± 67.9	624.6 ± 66.3	1339.9 ± 105.2	−183	0.41
SMb20980	79.9 ± 24.1	41 ± 18.4	101.3 ± 7.4	91.4 ± 12.9	−271	0.39
SMb21144	752 ± 50	777 ± 6.7	582 ± 35.1	624.9 ± 41.9	−441	0.36
SMb21154	478.8 ± 51.4	423.4 ± 15.4	403 ± 17.3	385.5 ± 35.5	−165	0.37
SMb21177 (*phoC*)*	1777.4 ± 98.3	51.6 ± 5.4	65.4 ± 8.4	60.5 ± 5.3	−85, −63	0.54, 0.48
SMb21192	1231 ± 47	2424.6 ± 122.1	1165.6 ± 104.3	2752.7 ± 88.2	−356	0.42
SMb21210	258.8 ± 14.6	278 ± 6	205.2 ± 7.9	278.3 ± 13.6	−120	0.36
SMb21216	53.7 ± 13.9	78.8 ± 28.7	85.7 ± 18.1	86 ± 15.7	−355	0.37
SMb21222	123.9 ± 16.4	172.3 ± 15.4	160 ± 14.2	194.4 ± 16.1	−352	0.37
SMb21270*	1974.1 32	861.1 ± 4.1	1253.1 ± 8.0	1343.7 ± 29.4	−91, −80	0.39, 0.37
SMb21307	36.4 ± 21.7	48.6 ± 15.7	39.2 ± 7.1	63.5 ± 12.2	−424	0.42
SMb21555	296.8 ± 22.4	343.7 ± 37.4	303.9 ± 2.2	463 ± 24.5	+165	0.39
SMc00009	1829 ± 2.5.6	3616.6 ± 329.7	1631.8 ± 107.7	2613.7 ± 157.8	−302	0.37
SMc00027	61.7 ± 24.9	48.6 ± 15.9	57.8 ± 14.1	97.9 ± 8.6	−368	0.39
SMc00042	604.5 ± 3.5	1059.1 ± 76.1	631 ± 38.4	1030.7 ± 40	−62	0.43
SMc00161	619.6 ± 37.8	909.3 ± 98.8	347.6 ± 16	522.1 ± 10.6	−228, −217	0.38, 0.59
SMc00171*	1530.9 ± 19.3	124.1 ± 10.6	94 ± 1.4	128.6 ± 3.4	−52	0.47
SMc00485	5771.6 ± 682.6	5146.4 ± 238.4	3506.4 ± 90.4	3809.5 ± 254.4	−196	0.39
SMc00618 (*ppk1*)*	6034 ± 272	366.8 ± 29.6	266.7 ± 33.5	255.2 ± 9.7	−47	0.46
SMc00801**	673.2 ± 92.9	1824.7 ± 163	1352.5 ± 120	1719 ± 153	−57	0.87
SMc00819 (*katA*)*	1415.4 ± 91.2	100 ± 12.3	111.8 ± 16.5	117.2 ± 14.6	−50	0.60
SMc00978	161.6 ± 5.7	117.1 ± 7.7	158.5 ± 12.9	106.9 ± 10.1	−48	0.59
SMc00982	61.2 ± 2.3	44.6 ± 6.4	71.5 ± 25.3	65.8 ± 5.1	−359	0.39
SMc01605*	1242.6 ± 4.4	270.2 ± 0.7	95.4 ± 3.6	88.6 ± 0.4	−58	0.52
SMc01635	80.5 ± 63.6	65.1 ± 16.2	69.4 ± 7.8	78.6 ± 5.3	−207	0.36
SMc01723*	894.3 ± 3.1	314.8 ± 1.2	190.0 ± 3.6	308.2 ± 7.3	−64	0.41
SMc01849	188.5 ± 10.9	171.9 ± 4.6	175.1 ± 39.7	188.2 ± 22.7	−321	0.42
SMc01852 (*pfk*)*	734.3 ± 126	288.9 ± 26.5	204.6 ± 9	217.6 ± 23.7	−105	0.49
SMc01907*	955.9 ± 24.3	186.8 ± 17.5	85.9 ± 2.8	125.2 ± 4.4	−82	0.72
SMc01934	406.5 ± 22.4	706.4 ± 44.4	440.7 ± 21.7	607.6 ± 26.9	−283	0.39
SMc01952	101.7 ± 10.2	100.6 ± 13.4	123.9 ± 10.9	118.6 ± 12.8	−299	0.37
SMc02146 (*pstS*)*	2658.6 ± 193	100.7 ± 23.4	81.3 ± 19.8	78.7 ± 12.4	−115, −104	0.72, 0.38
SMc02315	54 ± 5.6	49.5 ± 2.6	76.4 ± 16.8	56.5 ± 8.9	−134	0.44
SMc02601**	282.4 ± 19	818.9 ± 47.3	1036.6 ± 57	1426 ± 37.9	−99	0.50
SMc02634*	2315.4 ± 146	111.4 ± 9.8	90.4 ± 0.3	114.9 ± 4.2	−77	0.56
SMc02675	342.7 ± 9.4	308.9 ± 30.5	30.7 ± 0.9	44.6 ± 1.9	−410	0.78
SMc02689	977.9 ± 54	677.5 ± 116.7	1204.5 ± 86.4	717.7 ± 4.3	−106	0.5
SMc02862 (*orfA-pit*)**	46.2 ± 0.18	111.3 ± 3.9	99.8 ± 6.6	123.8 ± 6.4	−83	0.37
SMc02863 (*recF*)	42.3 ± 40.8	36.4 ± 12	71.8 ± 5	73.6 ± 12.5	−195	0.45
SMc02886	618.4 ± 12.8	731.9 ± 45.3	290.6 ± 24.6	347.7 ± 13.3	−90	0.51
SMc02976	398.6 ± 19.2	393 ± 47	233.3 ± 19.3	307.5 ± 15.3	−73	0.36
SMc03124*	4081.4 ± 13.2	130.5 ± 1.0	126.5 ± 2.6	109.7 ± 2.6	−93	0.44
SMc03174	137.6 ± 9	196.3 ± 14.7	140.4 ± 15.1	190.7 ± 29.1	−423	0.41
SMc03243(*phoA1*)*	1761.4 ± 159	108.4 ± 8.3	76 ± 5.3	109.6 ± 5.1	−64, −97	0.41, 0.37
SMc03823	153.3 ± 17.8	587.6 ± 26.5	218.9 ± 8.1	728.8 ± 45.7	−206	0.46
SMc03844	260.57 ± 1.2	438.1 ± 51.2	397.9 ± 43.3	557.4 ± 23.9	−33	0.42
SMc03975	367.3 ± 25.1	544.2 ± 67.3	704.6 ± 10.8	1085 ± 29.5	+22	0.56
SMc04053*	1048 ± 80.1	149.2 ± 18	125.6 ± 29.4	164.6 ± 17.7	−142	0.39
SMc04144	235.3 ± 8	247.6 ± 24.8	173.5 ± 25	220.7 ± 13.8	−46	0.39
SMc04213	161.3 ± 6.7	170 ± 1.4	185.8 ± 11.2	168.3 ± 20.1	−169	0.36
SMc04317 (*afuA*)*	3284.3 ± 299	50.6 ± 23.7	76.8 ± 9.3	59.7 ± 15.7	−51	0.58
SMc04458	26.4 ± 9.1	26.7 ± 1.4	52.4 ± 6.7	54.3 ± 7.9	−230	0.37
SMc04478	1498.2 ± 100.6	3333.8 ± 118.6	1632 ± 62.7	3952.4 ± 312.8	−52	0.57

[a]GusA transcriptional gene fusion and expression demonstrating altered gene expression in response to Pi conditions. β-glucuronidase activities are expressed as Miller units and corresponding alkaline Phosphatase activities (data not shown) were determined as described in Materials and Methods. Each value corresponds to the mean of triplicate assays with (±) standard error. Background β-glucuronidase values from negative control (*S.meliloti* strains with no promoter fusions) were in the range of 42–55 Miller units. SMc02862 (*orfA-pit*) was a *lacZ* fusion.

[b]Gene fusions integrated into the genome are shown in grey, others were generated in the plasmid pFUS1.

[c]*PhoB-dependent up-regulated genes, **PhoB-dependent down-regulated genes.

The Pho boxes from genes in '**bold**' were also part of the weight matrix (Table 1).

smc00161 is annotated to encode an NH_3-dependent NAD^+ synthetase and the promoter region of this genes was predicted to carry a promoter Pho box (Supplementary Table S1).

Comparison of Pho box predictions with DNA microarray data

Employing DNA microarrays, Krol and Becker (31) identified 98 genes (some of which were in operons) that were more than 3-fold induced in a *phoB*-dependent manner upon Pi limitation. An additional 50 genes showed a strong increase in expression under phosphate limitation in a partially *phoB*-dependent or *phoB*-independent manner. Krol and Becker (31) also identified potential Pho-box sequences with ⩽2 mismatches from the Pho-box consensus sequence TG(A/T)CA (C/A)-NNNN-C(C/T)(G/T)TCA(C/T) defined by Summers *et al*. (16). Of the 19 Pho-box promoters identified by Krol and Becker (31), 14 were also identified with our weight matrix (Table 3) and data from our gene fusion experiments (Table 2) revealed that 13 of these 14 genes were regulated by media Pi in a PhoB dependent manner (Table 2). A reporter fusion to the remaining gene, *sma0612* has yet to be examined. Of the five Pho-boxes not identified by our weight matrix, four were unusual (*sma1809*, *sma1822*, s*mc00170* (*sinR*) and *smc00429*) as they contained 3 or 5 nt in the region flanked by the 7 nt direct repeats instead of the 4 (see Table 1). The Pho-box upstream of the remaining gene, *sma0045*, lies on the opposite strand to *sma0045* and thus would not be included in our predictions. However reporter gene fusions to these genes should be analyzed as the microarray experiments suggested these genes were induced in a *phoB*-dependent manner (31).

In addition to the 13 Pho regulon members predicted both here and by Krol and Becker (31), both the weight matrix data and data from reporter fusion assays identified an additional 10 genes whose expression was PhoB-regulated in response to Pi limitation (Tables 2 and 3). With the exceptions of *smb20843* (*algI*), *smc00618*(*ppk*), *smc02601*(*nadA*) and *smc00801*(hypothetical, global homology), these genes

Table 3. Pho regulon members in *S.meliloti* as identified by this study and the microarray analysis of Krol and Becker (31).

No.	Gene	Study	Pho box motif	[d]Dist	Score	Description	*Genomes
1	SMb20759	Known	ATGTCACAAGCCTGTCAT	−66	0.62	*phnG*, carbon-phosphorus lyase component	*4
2	SMb21177	Known	CTGTTACAGAACCTACAC	−85	0.54	*phoC*, phosphate ABC transporter ATP binding	*6
		Known	CTGACACTGCGCTTTCAT	−63	0.48		
3	SMc02862	Known	CTGTGGGAAAGCCGTTTT	−83	0.37	*orfA-pit*, inorganic phosphate transporter	*4
4	SMa0612	Overlap[a]	TTGTCAGGGGCCTTTCAT	−18	0.36	*FixN3*, cytochrome *c* oxidase subunit 1	
5	Smc00171	Overlap[a,b]	ATGTCATCTTCCTGAAAC	−52	0.47	Conserved hypothetical protein	*5
6	Smc01605	Overlap[a,b]	CTGTCACATAGGCTTCAT	−58	0.52	Putative ABC transporter periplasmic binding	*6
7	Smc01907	Overlap[a,b]	CTGTCATCCAGCTGTTAC	−82	0.72	*phoA2*, putative alkaline phosphatase	* 3
8	Smc02146	Overlap[a,b]	GTTTCACCGAACTGACAT	−104	0.38	*pstS*, phosphate-binding periplasmic protein	*11
		Overlap[a,b]	CTGTCATAAATGTTTCAC	−115	0.72		
9	SMc02174	Overlap[a]	TTGTGAACGATCTGTCGC	−76	0.36	Hypothetical transmembrane protein	*1
10	Smc02634	Overlap[a,b]	TTGTCGCAGAGCTGTCAT	−77	0.56	*phoA3*, putative alkaline phosphatase	*2
11	Smc03124	Overlap[a,b]	CTGTCACGAAAGTTTCAC	−93	0.44	Putative ABC transporter periplasmic binding	*5
12	SMc03243	Overlap[a,b]	CTGTAAAGCCGCTGACAT	−64	0.41	*phoA1*, putative alkaline phosphatase	*5
		Overlap[a,b]	CTGTCGCCAAACCGTCCG	−97	0.37		
13	SMc03961	Overlap[a]	TTGTCACAATCCTATCCT	−40	0.36	*sqdB*, sulfolipid biosynthesis protein	
14	SMc04317	Overlap[a,b]	CTGTCACATCCCTGTAAA	−51	0.58	*afuA*, iron-binding periplasmic protein	*3
15	SMb20843	Gene fusion[b,c]	ATTTTAAATCGCCGTCAC	−340	0.38	*algI*, involved in acetylating surface saccharide	
16	SMb21270	Gene fusion[b,c]	CTGGCATCTTGATGTAAT	−91	0.38	Putative transcriptional regulator	
17	SMc00618	Gene fusion[b]	CTGTCACAGTTCCGTCGT	−47	0.46	*ppk*, putative polyphosphate kinase protein	*10
18	SMc00801	Gene fusion[b]	TTGTCATAAAACTGTCAC	−57	0.87	Hypothetical transmembrane protein	
19	SMc01723	Gene fusion[b]	CCGTCACTACCCTGACAT	−64	0.41	Hypothetical transmembrane global homology	
20	SMc01852	Gene fusion[b]	CTGTCATTTAATTTCCAT	−105	0.49	*pfk*, pyrophosphate-fructose 6-Pi 1-phosphotransferase	
21	SMc02601	Gene fusion[b]	CTGTCACAAAACTTTCTA	−99	0.5	*nadA*, putative quinolinate synthetase A protein	
22	SMc04053	Gene fusion[b]	TTGACATAAAATAGTAAT	−142	0.39	Hypothetical protein	
23	SMb20876	Gene fusion[b,c]	AATTCGTCAATCTGTCAC	−98	0.43	Hypothetical protein	*1
24	SMc00819	Gene fusion[b,c]	CTGTCGTTCAGCCGTCAC	−50	0.6	*katA*, catalase hydroperoxidase HPII	*4
25	SMa1961	Microarray[c]	CTGACAAACCGCCTTAGC	−128	0.37	Putative polyhydroxyalkanoate depolymerase	
26	SMb20037	Microarray[c]	ATGACGCTCACCCGTCAT	−389	0.39	*aroE2*, putative shikimate 5-dehydrogenase	
27	SMb21317	Microarray[c]	CTGTCATGCACCTGCATC	−385	0.39	*expG*, regulator of exopolysaccharide II synthesis	*1
28	SMc00620	Microarray[c]	CTTTCACCTCGCTGTAAG	−70	0.41	Hypothetical signal peptide protein	* 2
29	SMc00772	Microarray[c]	AGGTCGTCAAGGCGTCAT	−103	0.36	*potH*, probable putrescine transport permease	*1
30	SMc02490	Microarray[c]	TTGTCATTGTTCGATCAT	−41	0.38	Hypothetical protein	
31	SMc03994	Microarray[c]	CTGTCAGACAGGTTCAAT	−47	0.37	*suhB*, putative inositol monophosphatase	
32	SMc04239	Microarray[c]	CTGTCGTGAAGGCTTCAT	−98	0.4	Hypothetical protein	
33	SMc04280	Microarray[c]	CTTTTGTAAAGATTTCAT	−81	0.37	Hypothetical signal peptide protein	
34	SMc01848	Microarray[c]	TCGTCATCAAAGTGTAGC	−47	0.41	Hypothetical protein (*btaA*-like)	* 2

[a]Pho-box sequences detected in this study and also by Krol and Becker (31).
[b]Pho-box sequences detected in this study and gene fusions tested for PhoB-dependent regulation.
[c]Pho-box sequences detected only by our matrix and PhoB-dependent regulation shown by microarray data [Krol and Becker (31)]. No gene fusions were tested for #25–34.
[d]'Dist' indicates the relative distance from the annotated translational start sites.
*Indicates the conserved Pho regulon homologs having predicted Pho box in other scanned genomes (see Tables 5–8).

also showed PhoB-dependent transcription in microarray studies. The failure to detect repression of *smc02601*(*nadA*) and *smc00801*expression in microarray experiments is not surprising as the microarray experiments also failed to detect *orfA-pit* repression and this operon is known to be repressed by PhoB (25). The failure to detect induction of *smb20843* (*algI*), *smc00618* (*ppk*), upon Pi-starvation is more surprising as these genes appear to be highly regulated in the gene fusion experiments. Moreover expression of *ppk* is known to be Pi-starvation induced in many organisms. The differences between the microarray and gene fusion data could result from several factors including differences in experimental growth conditions as in microarray experiments cells were grown in 100 μM Pi source as the Pi-limitation condition. Alternatively, it is possible that the particular probes employed for *ppk* and *smb20843* yielded low signals. Through our weight matrix scan, Pho-box sequences were also found upstream of three more genes, *sma2410* (*rhbF*), *smc01296* (*rpsN*) and *smc01820* (putative N-carbamyl-L-amino acid aminohydrolase) (Supplementary Table S1). Promoter fusions to these three genes have not been tested yet. These genes however are shown to be repressed in a PhoB-independent manner in microarray studies (31). Further studies are required to analyze the regulation of these genes and the nature of their associated Pho-box sequences.

We note that in the case of *orfA-pit*, the Pho-box identified by Krol and Becker (31) lay on the opposite strand to the *orfA-pit* genes and is different from that identified by Bardin *et al*. (28). Since *orfA-pit* expression is negatively regulated by PhoB, it is of interest to determine the actual PhoB binding site as little is known regarding how PhoB represses transcription. In summary, of 96 genes with upstream Pho-boxes predicted by the frequency matrix genome analysis, 34 appear to be Pi and PhoB regulated as revealed from gene fusion and microarray analysis data (Table 3).

Analysis of predicted Pho regulon members across proteobacterial genomes

It is reasonable to assume that at least part of the physiological response to Pi-limitation will be conserved. As the Pho-box sequence identified by the PhoB proteins of different organisms appears to be conserved (14,16,24,28,42–44), we used the Pho-box frequency matrix described above (Table 1) to search the genomes of twelve gram negative bacteria (Table 4) for PhoB-binding sites using the same criteria as employed for *S.meliloti*. Genes that lay downstream of a predicted Pho-box with scores greater than 0.35 were further examined. We identified genes, such as *pstSCAB*, *phoA*, *ugpA*, *phn* and *ppk* that are known to be associated with phosphate metabolism (Tables 5–7). The *pstS* gene encodes the Pi-binding protein of the high affinity PstSCAB transport system (18,27) and expression of this system in *E.coli*, *S.meliloti* and *Pseudomonas aeruginosa* is known to be highly induced under Pi-limiting conditions and is PhoB dependent (27). In a number of organisms, such as *Caulobacter crescentus*, the *pstS* gene transcript is separate from the *pstCAB-phoUB* transcript and in these cases predicted Pho-boxes are also located upstream of the *pstC* gene (see Table 5).

It was striking that multiple18 bp Pho-box sequences were predicted upstream of the *pst* genes in all of the genomes examined (Table 5). Multiple Pho-boxes consisted of overlapping 7 bp direct repeats separated by 4 bp spacers. The frequency matrix detected consecutive 18 bp elements and adding a terminal 4 bp spacer formed consecutive 22 bp PhoB binding sites as defined by Blanco *et al*. (23). The two 11 bp direct repeat sequences bind the PhoB monomers head to tail (23). The *pstS* promoters from *E.coli* K12 and O157:H7 are predicted to contain five and six of these 11 bp direct repeats, respectively. The large number of Pho-boxes in all of the *pstS* promoter regions presumably reflects the importance of the PstSCAB high affinity transport system in the uptake of Pi under Pi-limiting conditions. Other genes associated with phosphate metabolism for which multiple Pho-boxes sequences were detected included alkaline phosphatase-like proteins (*phoA*), genes involved in phosphonate uptake and metabolism (*phn*), in glycerol-3-phosphate uptake (*ugp* and *glp*), the regulatory genes *phoB* and *phoR* (Table 5) and genes encoding polyphosphate kinase (Table 6).

In addition to the previously reported Pho-box in the *orfA-pit* promoter region of *S.meliloti*, Pho-box sequences were also detected in the promoter region of the *orfA-pit* orthologues in the α-proteobacteria, *Bradyrhizobium japonicum* and *Mesorhizobium loti* and the γ-proteobacteria *Pseudomonas putida* and *Acinetobacter* sp (Table 5). Bardin *et al*. (28) showed that the expression of *orfA-pit* in *S.meliloti* is repressed upon Pi-starvation, unlike in *E.coli* where the *pit*

Table 4. List of proteobacterial genomes scanned for the presence of Pho-boxes

No.	Bacterial genome	Accession no.	Number of predicted Pho regulons
1	*Acinetobacter* sp. ADP1	NC_005966	56
2	*A.tumefaciens* C58 Chromosome circular	NC_003304	51
	A.tumefaciens C58 Chromosome linear	NC_003305	35
	A.tumefaciens C58 Plasmid AT	NC_003306	7
	A.tumefaciens C58 Plasmid Ti	NC_003308	6
3	*B.japonicum* USDA 110	NC_004463	87
4	*B.melitensis* 16 M chromosome I	NC_003317	25
	B.melitensis 16 M chromosome II	NC_003318	33
5	*B.suis* 1330 chromosome I	NC_004310	27
	B.suis 1330 chromosome II	NC_004311	32
6	*C.crescentus* CB15	NC_002696	62
7	*E.coli* K12	NC_000913	107
8	*E.coli* O157:H7 Chromosome	NC_002695	96
	E.coli O157:H7 Plasmid pO157	NC_002128	5
	E.coli O157:H7 Plasmid pOSAK1	NC_002127	2
9	*M.loti* MAFF303099 for chromosome	NC_002678	71
	M.loti MAFF303099 plasmid pMLa	NC_002679	16
	M.loti MAFF303099 plasmid pMLb	NC_002682	7
10	*P.aeruginosa* PAO1	NC_002516	73
11	*P.putida* KT2440	NC_002947	71
12	*S.meliloti* 1021 chromosome	NC_003047	56
	S.meliloti 1021 plasmid pSymA	NC_003037	16
	S.meliloti 1021 plasmid pSymB	NC_003078	24
13	*Rhizobium* sp. NGR234 plasmid pNGR234a	NC_000914	17
	Rhizobium sp. NGR234 megaplasmid 2 contig 1	AY316747	4
	Rhizobium sp. NGR234 megaplasmid 2 contig 2	AY316746	3

Table 5. Predicted Pho-regulon orthologue groups related to Pi metabolism

Group	Gene	Pho box motif	Dist	Score	Genome source	Description-predicted function
A	**Phosphate transport (high affinity)**					
1	SMc02146	CTGTCATAAATGTTTCAC	−115	0.72	*S.meliloti*	*pstS*, putative periplasmic phosphate-binding protein
		GTTTCACCGAACTGACAT	−104	0.38		
2	ACIAD0279	CTGTCATTAAAGTTTCAT	−64	0.71	*Acinetobacter sp.*ADP1	*phoU*, transcriptional repressor for high affinity phosphate uptake
3	ACIAD1212	GCGTCATTAATTTGTCAC	−98	0.38	*Acinetobacter sp.* ADP1	*pstS*, putative ABC phosphate transporter (periplasmic-binding)
		TTGTCACTGAATTGTCAT	−87	0.54		
4	Atu0421	TTGACATTTCCCATTCAT	−151	0.37	*A.tumefaciens*	*pstC*, ABC transporter, membrane spanning protein
5	Atu0420	TTGTCACAAATCTTTCGT	−117	0.54	*A.tumefaciens*	*pstS*, ABC transporter, substrate binding protein [phosphate]
		CTTTCGTCAAAGTGACAT	−106	0.36		
6	blr109	CTGTCATCCGACTGTCAC	−88	0.71	*B.japonicum*	*pstS,* ABC transporter phosphate-binding protein
		CTGTCACGAAACCTTCGT	−77	0.48		
7	BMEI1989	GTGTCATATAAGTGTAAT	−76	0.56	*B.melitensis*	*pstS,* putative periplasmic phosphate-binding protein
		CTGTAATATTCGTGTCAT	−87	0.53		
		TTGCAACAAACCTGTAAT	−98	0.37		
8	BR2138	TTGCAACAAACCTGTAAT	−89	0.36	*B.suis*	*pstS,* phosphate ABC transporter, phosphate-binding protein
		CTGTAATATTAGTGTCAT	−78	0.56		
		GTGTCATATGAGTGTAAT	−67	0.41		
9	CC0290	TTGTCGTCAAACTGTCAT	−104	0.68	*C.crescentus*	*pstCAB-phoUB*, phosphate ABC transporter, permease protein
		CTGTCATGTAATTGTCGC	−93	0.48		
10	CC1515	CTGTCGTCAAACTGTGAC	−62	0.57	*C.crescentus*	*pstS,* putative ABC transporter, periplasmic phosphate-binding
		CTGCCTTAAAACTGTCGT	−73	0.38		
11	b3728	CTGTCACCTGTTTGTCCT	−56	0.36	*E.coli* K12	*pstS,* high-affinity phosphate transport phosphate-binding protein
		CTTACATATAACTGTCAC	−67	0.62		
		CTGTCATATTCCTTACAT	−78	0.64		
		CTGTCATAAAACTGTCAT	−89	0.97		
12	ECs4664	CTGTCACCTGTTTGTCCT	−56	0.36	*E.coli* O157:H7	*pstS*, periplasmic phosphate-binding protein
		CTTACATATAACTGTCAC	−67	0.63		
		CTGTCACATTCCTTACAT	−78	0.62		
		GTGTCATCAAACTGTCAC	−89	0.67		
		CTTTCCTTTGGGTGTCAT	−100	0.37		
13	mll3723	CTGTCATGCGACTGTAAT	−79	0.58	*M.loti*	*pstS*, periplasmic phosphate-binding protein, (PBP)
		CTGACACGAAACTGTCAT	−90	0.65		
14	PP5328	TTGTAATGTTTTTGTCAC	−302	0.38	*P.putida*	*pstC*, phosphate ABC transporter, permease protein
15	PP5329	TTGTCACAATGAAGTCAT	−75	0.41	*P.putida*	*pstS,* ABC transporter, periplasmic phosphate-binding protein
		CTTTCATCCAATTGTCAC	−86	0.56		
16	PP2656	CTGCCATTCAATAGTCAC	−42	0.37	*P.putida*	*pstS*, phosphate ABC transporter, periplasmic phosphate-binding
		TAGTCACAAAGTTGTAAC	−31	0.49		
		TTGTAACACAGCTGATGC	−20	0.36		
17	PA5368	CTGTCATCTGTCTGTCAT	−277	0.74	*P.aeruginosa*	*pstC*, membrane protein component of ABC phosphate transporter
		GCGTCATGTTGCTGTCAT	−288	0.39		
18	PA5369	CTTTCATAGAGTCTTCAT	−60	0.53	*P.aeruginosa*	*pstS*-like, hypothetical protein
		CTGTCATATTCCTTTCAT	−71	0.78		
		ATTTCATCCAACTGTCAT	−82	0.56		
B	**Phosphate transport (low affinity)**					
1	SMc02862	CTGTGGGAAAGCCGTTTT	−83	0.37	*S.meliloti*	*orfA-pit*, phosphate transport transmembrane protein
2	ACIAD1047	TTGTCATATAATTGTCAT	−51	0.73	*Acinetobacter sp.* ADP1	*pit*, phosphate transporter
3	bll3022	TTGTCATCCAGCCTTCAA	−49	0.53	*B.japonicum*	*pit*-like, low-affinity inorganic phosphate transporter
4	mll3637	TCGTCATACAGGGTTCAT	−56	0.36	*M.loti*	*pit*, phosphate transporter
5	PP1373	CTGTCATCTGCCTGTTAC	−67	0.56	*P.putida*	Low-affinity inorganic phosphate transporter
C	**Phosphonate / Phosphate metabolism**					
1	SMb21177	CTGACACTGCGCTTTCAT	−63	0.54	*S.meliloti*	*phoC*, phosphate uptake ABC transporter ATP-binding protein
		CTGTTACAGAACCTACAC	−85	0.48		
2	SMb20759	ATGTCACAAGCCTGTCAT	−66	0.4	*S.meliloti*	*phnG*, putative C-P (carbon-phosphorus lyase component protein
3	ACIAD0719	TTGATATCAAGCTTCCAT	−71	0.38	*Acinetobacter sp.* ADP1	*phnA*, putative alkylphosphonate uptake protein (PhnA)
4	Atu0181	ATGTCACACGCTTGTCAT	−67	0.47	*A.tumefaciens*	*phnG*, hypothetical protein
5	Atu0174	CTGACACTCCCGCTTCAC	−48	0.36	*A.tumefaciens*	*phnC*, ABC transporter, nucleotide binding/ ATPase [phosphonate]
		TCTTCATCAAACTGACAC	−59	0.37		
		ATGTAATATTTTCTTCAT	−70	0.36		
6	Atu6108	CTGTCAGAACACCCTGAT	−81	0.37	*A.tumefaciens*	*phnA*, alkylphosphonate uptake protein
7	blr7947	CTGTCATCGCTGCGTCAT	−99	0.62	*B.japonicum*	*phoC*, phosphonate metabolism protein
8	blr1221	TTGTCACGAAACAGTCAT	−12	0.47	*B.japonicum*	*phnG*, phosphonate metabolism protein
9	bll7947	ATGTCACACAAGTGTCAA	−44	0.52	*B.japonicum*	*phoC*, phosphonates transport system nucleotide binding/ATPase
10	CC0361	CCGTCACAAATCCGTTGC	−68	0.36	*C.crescentus*	Phosphonates ABC transporter, ATP-binding protein
11	ECs5088	CTGTTAGTCACTTTTAAT	−60	0.43	*E.coli* O157:H7	*phnC*, ATP-binding component of phosphonate transport
12	b4106	CTGTTAGTCACTTTTAAT	−60	0.42	*E.coli* K12	*phnC*, ATP-binding component of phosphonate transport
13	mll9156	TAGTCATCTCACTGTCAT	−315	0.56	*M.loti*	*phnM* homolog, phosphonate metabolism protein

Table 5. *Continued*

Group	Gene	Pho box motif	Dist	Score	Genome source	Description-predicted function
14	mlr3342	CTGTCATCCACCCGTCAT	−80	0.73	*M.loti*	*phnG*, similar to phnG gene product (phosphonate metabolism)
15	PP2209	CTGACTTATAGCTGGGAT	−85	0.37	*P.putida*	*phnW*, 2-aminoethylphosphonate:pyruvate aminotransferase
16	PA3384	CTGTCATCGTCACTTCAC	−84	0.44	*P.aeruginosa*	*phnC*, ATP-binding component of ABC phosphonate transporter
		TTGTCACTCGACTGTCAT	−95	0.59		
D	**Pho response regulator**					
1	ACIAD3557	TTGCAATGATACTGTCAT	−74	0.37	*Acinetobacter sp.* ADP1	*phoB-phoR*, positive response regulator for the pho regulon PhoR
		TTGTAATAAGCCTGTCAT	−47	0.49	*Acinetobacter sp.* ADP1	
2	Atu0419	ATGATGAATCTCTGTCAT	−58	0.36	*A.tumefaciens*	*phoR*, two component sensor kinase
		CTGTCATGAAGCTGGCCT	−47	0.38		
3	blr1090	CTGTAATCTTGCTGACGT	−179	0.37	*B.japonicum*	*phoR*, phosphate regulon, two-component sensor histidine kinase
4	BMEI1624	CTGTCTTTGAACTGTCAC	−219	0.68	*B.melitensis*	*phoR*, phosphate regulon sensor protein
5	BR0298	CTGTCTTTGAACTGTCAC	−137	0.68	*B.suis*	*phoR*, putative sensor histidine kinase
6	b0399	TTTTCATAAATCTGTCAT	−64	0.41	*E.coli* K12	*phoB*, positive response regulator for pho regulon, sensor is PhoR
		CTGTCATAAATCTGACGC	−53	0.83		
7	ECs0449	TTTTCATAAATCTGTCAT	−64	0.72	*E.coli* O157:H7	*phoB-phoR* operon, positive response regulator for pho regulon
		CTGTCATAAATCTGACGC	−53	0.65		
8	PP5320	CTGTCACACAGCTGCAAT	−15	0.68	*P.putida*	*phoB-phoR*, DNA-binding response regulator
		CTGCAATAATTCCGTTAT	−4	0.39		
9	PA5360	GTGTCACATACCTGACAC	−50	0.54	*P.aeruginosa*	*phoB-phoR*, two-component response regulator PhoB
		CTGACACAATTTCGTTAT	−39	0.42		
E	**Glycerol 3-phosphate transport**					
1	ACIAD1317	CAGTCATTGAATCTTCAT	−168	0.42	*Acinetobacter sp.* ADP1	*gpsA*, glycerol-3-phosphate dehydrogenase, biosynthetic
2	Atu5058	CTGTCATCAAAACGTCGC	−75	0.47	*A.tumefaciens*	*ugpB*, ABC transporter, substrate binding protein
		TTGCTAGACAGCCGTCAT	−29	0.37		
3	Atu0305	GTGTAATAAATCTGACAC	−76	0.44	*A.tumefaciens*	*ugpA*, ABC transporter, substrate binding protein
		CTGACACGGAACTGTCAA	−87	0.47		
		CTGTCAAAGAGGCGTCAT	−98	0.63		
4	bll0733	CAGTCATGTGAACGTCAT	−55	0.36	*B.japonicum*	ABC transporter glycerol-3-phosphate-binding protein
5	blr2436	CCTTCGCCTCTCTTTCAT	−76	0.37	*B.japonicum*	*glpD*, glycerol-3-phosphate dehydrogenase
		CTTTCATCTTGCTTTCGC	−65	0.39		
6	b3453	AAGTTATTTTTCTGTAAT	−75	0.37	*E.coli* K12	*ugpB*, sn-glycerol 3-phosphate transport periplasmic binding protein
		CTATCTTACAAATGTAAC	−97	0.39		
		TTGTCATCTTTCTGACAC	−119	0.41		
7	ECs4299	AAGTTATTTTTCTGTAAT	−75	0.37	*E.coli* O157:H7	*ugpB*, sn-glycerol 3-phosphate transport periplasmic binding protein
		CTATCTTACAAATGTAAC	−97	0.39		
		CTGACACCTTACTATCTT	−108	0.41		
		CCGTCACCGCCTTGTCAT	−130	0.4		
8	mll3503	CTGTCACATACCTTCTCT	−61	0.37	*M.loti*	*ugpB*, sn-glycerol 3-phosphate transport system; periplasmic binding
F	**Phosphatases**					
1	SMc03243	CTGTAAAGCCGCTGACAT	−64	0.41	*S.meliloti*	*phoA*, putative alkaline phosphatase
		CTGTCGCCAAACCGTCCG	−97	0.37		
2	Smc01907	CTGTCATCCAGCTGTTAC	−82	0.72	*S.meliloti*	Hypothetical transmembrane protein
3	Smc02634	TTGTCGCAGAGCTGTCAT	−77	0.56	*S.meliloti*	Hypothetical protein, phosphatase
4	Atu1263	ATGTCATGCCACTGTCAC	−51	0.59	*A.tumefaciens*	Conserved hypothetical protein
5	BMEII0655	CCGTCATTCCTGTGTAAT	−53	0.38	*B.melitensis*	*phoA*, alkaline phosphatase
6	BR1200	CTGTCACACGCCTGCAAT	−80	0.44	*B.suis*	*phoA*, alkaline phosphatase
7	BRA0616	CTGTCATCGTTTCTTCGT	−75	0.39	*B.suis*	Hypothetical protein, *phoA* like gene
		CCGTCATTCCTGTGTAAT	−53	0.37		
8	b0383	CTGTCATAAAGTTGTCAC	−63	0.88	*E.coli* K12	*phoA*, alkaline phosphatase
9	ECs0433	CTTTTCAACAGCTGTCAT	−5	0.39	*E.coli* O157:H7	*phoA*, alkaline phosphatase
		CTGTCATAAAGTTGTCAC	+6	0.86		
10	ECs4053	TTGTCAGGTATCTGTATC	−361	0.37	*E.coli* O157:H7	*phoA*, putative alkaline phosphatase I
11	mll2704	GTGTCACATGGCCGTCAC	−45	0.46	*M.loti*	Probable acid phosphatase
12	mll4115	CTGTTATCAAAACTTCAT	−73	0.52	*M.loti*	*phoA*, secreted alkaline phosphatase
13	PP1044	CTGCCATCAAACTGTAAT	−45	0.66	*P.putida*	(*uxpA*) lipoprotein UxpA
		CTGTCGGATTTCTGCCAT	−56	0.41		
14	PA3296	TTGTCACAAGCCCGTCAT	−82	0.53	*P.aeruginosa*	*phoA*, alkaline phosphatase
15	PA2635	CTGTCATCGTCCCGTCGC	−53	0.39	*P.aeruginosa*	Hypothetical protein
G	**Membrane lipids**					
1	SMc01848	TCGTCATCAAAGTGTAGC	−47	0.41	*S.meliloti*	Hypothetical protein (*btaA*-like)
2	Atu2119	CTGTCATCAAACTGTAGC	−44	0.58	*A.tumefaciens*	Hypothetical protein (*btaA*-like)
3	mlr1574	CTGTCACCGGCCTGTCAT	+1	0.55	*M.loti*	Hypothetical protein (*btaA*-like)
H	**Exopolysaccharide**					
1	SMb21317	CTGTCATGCACCTGCATC	−385	0.39	*S.meliloti*	*expG*, activator of exopolysaccharide II synthesis
2	SMc02851	CTTTCAAAGAGCCGCCAC	−158	0.37	*S.meliloti*	Putative transcription regulator protein
3	bll5036	TTGTGGCACAGGCTTAAT	−145	0.36	*B.japonicum*	*mucS*, transcriptional regulatory protein

'Dist' indicates the relative distance from the annotated translational start sites.

Table 6. Conserved *ppk* Pho box motifs in various proteobacteria

No.	Gene	Source/Pho box motif	Gene annotation/Distance from ATG[a]	Score
1	ACIAD1782	*Acinetobacter sp.* ADP1	*ppk* polyphosphate kinase	
		TTGTAACTCGTTTGTAAC	−107	0.36
2	Atu1144	*A.tumefaciens*	*ppk* polyphosphate kinase	
		CTGTCACAGTTCCGTCGT	−84	0.46
		CCGTCGTCAAACTGTTAT	−73	0.39
3	bll2813	*B.japonicum*	*ppk2* hypothetical protein	
		CTGCCATCGCCGTGTCAA	−365	0.36
4	bll4122	*B.japonicum*	*ppk* polyphosphate kinase	
		ATGTCATCGAAACGTCAT	−59	0.48
5	BMEI1205	*B.melitensis*	*ppk* polyphosphate kinase,	
		TTGTCATATGACAGCCAT	+10	0.37
		TTGCTATCAAATTGTCAT	−1	0.39
6	BR0748	*B.suis*	*ppk* polyphosphate kinase	
		TTGTCATATGACAGCCAT	−101	0.37
		TTGCTATCAAATTGTCAT	−112	0.4
7	CC1710	*C.crescentus*	*ppk* polyphosphate kinase,	
		CTGTCTTCACCGCGTCAT	+106	0.37
8	mlr8161	*M.loti*	*ppk* polyphosphate kinase	
		ATGTTAGAGCACCGCCAT	+18	0.36
9	mlr8387	*M.loti*	*ppk2* hypothetical protein	
		CTGCAATAAAACCGTCAC	+74	0.51
10	PA5243	*P.aeruginosa*	*ppk* delta-aminolevulinic acid dehydratase	
		TTGTTGCCAATCCGTCAT	−56	0.39
11	PA2428	*P.aeruginosa*	*ppk2* hypothetical protein	
		CCGTCACCAAACCGTCAT	−52	0.53
		TTTTCATCTAACCGTCAC	−63	0.51
12	PP5216	*P.putida*	*ppk* exopolyphosphatase	
		CTGTCATATGGCCGTCAT	−119	0.73
13	PP5217	*P.putida*	*ppk* polyphosphate kinase	
		GTGTAACACGGGCGTCAT	−122	0.36
14	PP1752	*P.putida*	*ppk2* hypothetical protein	
		CTGTGGGAGCGCGTTCAT	−70	0.37
15	SMc00618	*S.meliloti*	*ppk* putative polyphosphate kinase	
		CTGTCACAGTTCCGTCGT	−47	0.46

[a]Indicates the relative distance from the annotated translational start sites.

genes appear to be constitutively expressed (45) and for which no Pho-boxes were detected. The identification of putative Pho-boxes upstream of the *orfA-pit* genes in other bacteria suggests that these may also be repressed by Pi-starvation and that such repression may be a widespread phenomenon.

A number of predicted Pho regulon members not normally associated with Pi metabolism were identified in several genomes (Table 7). One of the genes in this category was *katA* encoding catalase and was recently shown to be PhoB dependent in *S.meliloti* and *P.aeruginosa* in Pi-starvation conditions (46). The detection of Pho-box elements upstream of the *katA* genes of *C.crescentus* and *P.putida* suggests that *katA* expression in these organisms is also PhoB regulated. Pho-boxes upstream of several *S.meliloti* ABC-class transport systems were also detected upstream of homologous clusters in other bacteria. These were *smc01605*, *smc04317* (*afuA*) and *smc03124* (Table 7). Both the *afuABC* and *smc01605* gene clusters in *S.meliloti* are annotated as putatively involved in Fe^{+3} transport, however definitive evidence is lacking. Choa *et al*., (47) did not find either of these ABC transport systems to be up-regulated in *S.meliloti* when grown in iron-limiting conditions. Therefore, it appears unlikely that they are actually involved in iron transport. A third ABC system in *S.meliloti*, *smc03124*, with conserved Pho-box sequences in other proteobacteria (Table 7), is annotated as a putative peptide binding protein. The actual substrate(s) transported by this system is unknown.

We identified a putative Pho-box upstream of *smc00772* (*potH*)- gene clusters well as orthologues in *M.loti* and *Brucella suis* (Table 7). Although fusion data for *smc00772* is unavailable, the *potFGHI* ABC-class, putative putrescine transporter cluster was identified as upregulated by Pi-limitation in the microarray analysis (31), although no Pho-box was identified by them. The putative Pho-box upstream of *smc00772* (*potH*) lies within the coding region of *potG* (*smc00771*), instead of upstream of the regulator (*potF*). The fact that Pho-box-like sequences were identified upstream of genes similar to *S.meliloti potH* in *M.loti* and *B.suis* suggest that putrescine transport may be PhoB-regulated across a range of organisms and should be further investigated.

In response to Pi-starvation, *S.meliloti* replaces phospholipids with other non-Pi-containing lipids sulphoquinovosyl diacylglycerols (SL), ornithine-containing lipids (OL) and diacylglyceryl-N,N,N-trimethylhomoserines (DGTS) (48,49). In *Rhodobacter sphaeroides* it was demonstrated that the *smc01848* homolog *btaA* is directly involved in DGTS biosynthesis (50) and recently Lopez-Lara *et al*. (51) established that *smc01848* and smc01849 (*btaA*B) are required for DGTS synthesis. A Pho-box is predicted 64 nt from the *smc01848* start codon and orthologs of *smc01848* in *M.loti* (*mlr1574*)

Table 7. Conserved Pho regulon members either of unknown and/or not clearly associated with Pi metabolism

Group	Gene	Pho box motif	Dist	Score	Genome source	Description:(predicted) function
A						
1	**SMc01605**	CTGTCACATAGGCTTCAT	−58	0.52	*S.meliloti*	ABC transporter, periplasmic substrate-binding Fe^{+3}
2	Atu2147	TTGCAATCCAACCGTCAC	−63	0.39	*A.tumefaciens*	ABC transporter, substrate binding protein
3	BMEII1120	CTGTCATCCAACCGTCAT	−96	0.73	*B.melitensis*	Iron(III)-binding periplasmic protein
4	BRA0115	CTGTCATCCAACCGTCAT	−79	0.77	*B.suis*	ABC transporter, periplasmic substrate-binding
5	mll3069	CTTTCATCTGGCTGTCAC	−58	0.55	*M.loti*	ABC transporter, periplasmic binding protein
6	PA3250	CTGACATGAAACCGTCAT	−124	0.58	*P.aeruginosa*	Hypothetical protein (smc01605 homolog)
7	PP1726	CCTACATCAAACTGTCAC	−119	0.39	*P.putida*	ABC transporter, periplasmic binding protein
B						
1	**SMc04317**	CTGTCACATCCCTGTAAA	−51	0.58	*S.meliloti*	*afuA*, iron-binding periplasmic protein
2	Atu0202	CTGGCATCCATCCGATAT	−66	0.37	*A.tumefaciens*	ABC transporter, substrate binding protein (iron)
3	afuA2	CTGTCACATACATGTTAT	−46	0.54	*A.tumefaciens*	Atu2014 ABC transporter, substrate binding
4	mll3626	CTGTCATCCACCAGTCAT	−79	0.62	*M.loti*	ABC transporter, periplasmic substrate-binding (iron)
5	y4fP	ATGTAATTAAACTGACAT	+24	0.53	*Rhizobium sp. NGR234*	Probable ABC transporter periplasmic binding
C						
1	**SMb20876**	AATTCGTCAATCTGTCAC	−98	0.43	*S.meliloti*	Hypothetical protein
2	Atu2329	CGGTCACACGCCTGTCAT	−161	0.51	*A.tumefaciens*	Conserved hypothetical protein
D						
1	**Smc00171**	ATGTCATCTTCCTGAAAC	−52	0.47	*S.meliloti*	Conserved hypothetical protein
2	Atu1649	CTGTCATGTTGCTGAAAC	−64	0.51	*A.tumefaciens*	Conserved hypothetical protein
3	bll5904	CGGTCATCCCCCTGTCAC	−24	0.52	*B.japonicum*	Hypothetical protein
4	CC3344	CTGTCAGAAAGTCCGCAT	−152	0.38	*C.crescentus*	Hypothetical protein
5	mll0806	CTGACATCGCGATTTCAC	−27	0.44	*M.loti*	Hypothetical protein
7	PP4510	TTGCCATGGCGCTGTCAT	−43	0.37	*P.putida*	Conserved hypothetical protein
E						
1	**SMc00819**	CTGTCGTTCAGCCGTCAC	−50	0.6	*S.meliloti*	*katA*, monofunctional catalase HPII
2	Atu4642	TAGTCATCTTCATGACAG	−133	0.37	*A.tumefaciens*	*katA*, bifunctional catalase/peroxidase HPI
3	CC3043	CGGTCGGTAAGGTGTCAC	−60	0.36	*C.crescentus*	Catalase/peroxidase
4	PA4236	CTGTCATTCATCCTTAAC	−157	0.67	*P.aeruginosa*	*katA*, monofunctional catalase HPII
5	PP3668	CTGAAAGAACACTGGCAT	−41	0.41	*P.putida*	Catalase/peroxidase HPI
F						
1	**SMc03124**	CTGTCACGAAAGTTTCAC	−93	0.44	*S.meliloti*	Putative periplasmic binding ABC transporter
2	Atu6138	TTGCAACAAAACTGTCAC	−77	0.43	*A.tumefaciens*	accA, ABC transporter, substrate binding protein
		CTGTCACCAAACTTTCAT	−66	0.76		
3	blr0308	CTGTCGGCTCACTGACAC	−445	0.45	*B.japonicum*	ABC transporter peptide-binding protein
4	BMEI1934	ATGTCATATTTCTGAAAT	−418	0.56	*B.melitensis*	Putative periplasmic binding ABC transporter
5	BMEII0505	CTGACACATGGCTGGAAC	−23	0.37	*B.melitensis*	Oligopeptide transport system permease protein
6	BRA0786	CTGACACATGGCTGGAAC	−23	0.37	*B.suis*	Peptide ABC transporter, substrate binding protein
7	mlr9269	CCCTCATAAAACCGTCAT	−85	0.47	*M.loti*	Extracellular solute-binding protein
8	mll9149	TTGTCATTCGACTGTCAT	−89	0.62	*M.loti*	Oligopeptide ABC transporter oligopeptide-binding
G						
	SMc00620	CTTTCACCTCGCTGTAAG	−70	0.41	*S.meliloti*	Hypothetical signal peptide protein
	blr7070	CTTTCACAAAACTGAAAC	−44	0.57	*B.japonicum*	Hypothetical protein
	mlr1518	CTGTCACGCAAGCATCAT	−47	0.38	*M.loti*	Hypothetical protein
H						
	SMc00772	AGGTCGTCAAGGCGTCAT	−103	0.36	*S.meliloti*	potH, putrescine transport system permease
	BR1610	TTGTCATCGCAGTGCCAT	−51	0.39	*B.suis*	Spermidine/putrescine ABC transporter, permease
I						
	SMc02174	TTGTGAACGATCTGTCGC	−76	0.36	*S.meliloti*	Hypothetical transmembrane protein
	mll8170	CTGTGAACAATGTGTCGC	−72	0.36	*M.loti*	Hypothetical protein

'Dist' indicates the relative distance from the annotated translational start sites.

and *Agrobacterium tumefaciens* (*atu2119*) also have predicted Pho boxes in the corresponding promoter regions (Table 5). These data strongly suggest that DGTS synthesis induced upon Pi limitation is mediated directly via PhoR-PhoB system.

Pi starvation and polyphosphate metabolism

Inorganic polyphosphates (polyPi) are linear polymers of orthophosphate residues linked by high-energy phosphoanhydride bonds. These polymers can vary in size from 3 to over 1000 phosphate residues. PolyPi is ubiquitous and the enzyme primarily responsible for polyPi synthesis in *E.coli* is polyP kinase (PPK), which uses the gamma phosphate of ATP to make the polymer. PolyPi can also be hydrolyzed to Pi either by exopolyhosphatases (PPX) or by endopolyphosphatases (PPN). The identification and assignment of Pho-boxes was sometimes complicated by differences in genome annotation, as in the case of genes encoding polyphosphate kinase (*ppk*) (Table 6). Here Pho boxes were predicted in the *ppk* promoter regions of 10 of the 12 genomes examined. However, the predicted Pho-box from both *M.loti* (52) and *C.crescentus* (53) were located within the annotated gene coding regions. Alignment of the Ppk amino acid sequence suggests that the actual start codons of the *ppk* genes in *M.loti* and *C.crescentus* are downstream

of the annotated start codons (data not shown). Our reporter gene fusion data showed that the *S.meliloti ppk* gene was strongly induced member of the Pho regulon. However, most strikingly, the weight matrix did not detect a *ppk* Pho box either from the *E.coli* K12 genome or the *E.coli* O157 genome, even at very low cut-off (0.18). In *E.coli* there is genetic evidence demonstrated that polyphosphates accumulate upon Pi starvation and depend on PhoB, although the *E.coli ppk* promoter has never been mapped. Therefore, it is likely that *E.coli* PhoB regulates *ppk* indirectly as suggested elsewhere (54).

CONCLUSION

Several complementary approaches were integrated to investigate the cellular response to Pi starvation. As a first step, computational identification of PhoB binding motifs predicted 96 potential Pho regulon members from the entire *S.meliloti* genome. These were subsequently investigated by genetic screening of transcriptional reporter gene fusions and through comparisons with recently available microarray data (31). It was found that 34 out of the 96 *in silico* predicted Pho regulon members were regulated by Pi concentration in a PhoB dependent manner (Table 3). These 34 Pho regulon members were analyzed *in silico* for conservation or co-occurrence across 12 genomes scanned (Tables 5 and 7). Nineteen of these 34 candidates were also predicted as having upstream Pho-boxes in at least one of the other genomes scanned in this study. The *in silico* analysis provided evidence for the conservation of a core Pho regulon in bacteria and suggests that these organisms share a common response to Pi limitation. Such a conservation is not surprising as for example in both plants and yeast one of the major responses to Pi-limitation is the induction of a high affinity Pi transport system and the induction of scavenging enzymes, such as alkaline phosphatases.

Extending the Pho-box analysis to many more genomes should define the core group of genes that respond to Pi-starvation. Further it will allow the identification of subgroups of genes, such as *katA*, whose expression is regulated by PhoB in some organisms but not in others. Analysis of the distribution of such data may lead to the recognition of associations between particular regulatory patterns and other phenotypic properties of the organisms.

SUPPLEMENTARY DATA

Supplemental Data are available at NAR online.

ACKNOWLEDGEMENTS

This work was supported with funding from the Natural Sciences and Engineering Research Council of Canada, from Genome Canada through the Ontario Genomics Institute and from the Ontario Research and Development Challenge Fund to T.M.F. The authors thank Dr Brain Golding for help in computing and he and Weilong Hao and Ying Fong for help and advice for the *in silico* comparison analysis. Funding to pay the Open Access publication charges for this article was provided by NSERC and Genome Canada.

Conflict of interest statement. None declared.

REFERENCES

1. McGuire,A.M. and Church,G.M. (2000) Predicting regulons and their *cis*-regulatory motifs by comparative genomics. *Nucleic Acids Res.*, **28**, 4523–4530.
2. Tompa,M., Li,N., Bailey,T.L., Church,G.M., De Moor,B., Eskin,E., Favorov,A.V., Frith,M.C., Fu,Y., Kent,W.J. *et al*. (2005) Assessing computational tools for the discovery of transcription factor binding sites. *Nat. Biotechnol.*, **23**, 137–144.
3. Thieffry,D., Salgado,H., Huerta,A.M. and Collado-Vides,J. (1998) Prediction of transcriptional regulatory sites in the complete genome sequence of *Escherichia coli* K12. *Bioinformatics*, **14**, 391–400.
4. Hertz,G.Z. and Stormo,G.D. (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics*, **15**, 563–577.
5. Stormo,G.D. (2000) DNA binding sites: representation and discovery. *Bioinformatics*, **16**, 16–23.
6. Fernandez De,Henestrosa,A.R., Ogi,T., Aoyagi,S., Chafin,D., Hayes,J.J., Ohmori,H. and Woodgate,R. (2000) Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Mol. Microbiol.*, **35**, 1560–1572.
7. Panina,E.M., Mironov,A.A. and Gelfand,M.S. (2001) Comparative analysis of FUR regulons in gamma-proteobacteria. *Nucleic Acids Res.*, **29**, 5195–5206.
8. Tan,K., Moreno-Hagelsieb,G., Collado-Vides,J. and Stormo,G.D. (2001) A comparative genomics approach to prediction of new members of regulons. *Genome Res.*, **11**, 566–584.
9. Baichoo,N., Wang,T., Ye,R. and Helmann,J.D. (2002) Global analysis of the *Bacillus subtilis* Fur regulon and the iron starvation stimulon. *Mol. Microbiol.*, **45**, 1613–1629.
10. Dombrecht,B., Marchal,K., Vanderleyden,J. and Michiels,J. (2002) Prediction and overview of the RpoN-regulon in closely related species of the *Rhizobiales*. *Genome Biol.*, **3**, research0076.1–research0076.11.
11. Gaballa,A., Wang,T., Ye,R.W. and Helmann,J.D. (2002) Functional analysis of the *Bacillus subtilis* Zur regulon. *J. Bacteriol.*, **184**, 6508–6514.
12. Zheng,D., Constantinidou,C., Hobman,J.L. and Minchin,S.D. (2004) Identification of the CRP regulon using *in vitro* and *in vivo* transcriptional profiling. *Nucleic Acids Res.*, **32**, 5874–5893.
13. Lee,T.Y., Makino,K., Shinagawa,H., Amemura,M. and Nakata,A. (1989) Phosphate regulon in members of the family *Enterobacteriaceae*: comparison of the PhoB-PhoR operons of *Escherichia coli, Shigella dysenteriae*, and *Klebsiella pneumoniae*. *J. Bacteriol.*, **171**, 6593–6599.
14. Wanner,B.L. (1993) Gene regulation by Phosphate in enteric bacteria. *J. Cell. Biochem.*, **51**, 47–54.
15. Scholten,M., Janssen,R., Bogaarts,C., van Strien,J. and Tommassen,J. (1995) The Pho regulon of *Shigella flexneri*. *Mol. Microbiol.*, **15**, 247–254.
16. Summers,M.L., Denton,M.C. and McDermott,T.R. (1999) Genes coding for Phosphotransacetylase and acetate kinase in *Sinorhizobium meliloti* are in an operon that is inducible by Phosphate stress and controlled by PhoB. *J. Bacteriol.*, **181**, 2217–2224.
17. Makino,K., Amemura,M., Kawamoto,T., Kimura,S., Shinagawa,H., Nakata,A. and Suzuki,M. (1996) DNA binding of PhoB and its interaction with RNA polymerase. *J. Mol. Biol.*, **259**, 15–26.
18. Kimura,S., Makino,K., Shinagawa,H., Amemura,M. and Nakata,A. (1989) Regulation of the Phosphate regulon of *Escherichia coli*: characterization of the promoter of the *pstS* gene. *Mol. Gen. Genet.*, **215**, 374–380.
19. Fiedler,U. and Weiss,V. (1995) A common switch in activation of the response regulators NtrC and PhoB: phosphorylation induces dimerization of the receiver modules. *EMBO J.*, **14**, 3696–3705.
20. Makino,K., Amemura,M., Kim,S.K., Nakata,A. and Shinagawa,H. (1993) Role of the sigma 70 subunit of RNA polymerase in transcriptional activation by activator protein PhoB in *Escherichia coli*. *Genes Dev.*, **7**, 149–160.

21. Sola,M., Gomis-Ruth,F.X., Serrano,L., Gonzalez,A. and Coll,M. (1999) Three-dimensional crystal structure of the transcription factor PhoB receiver domain. *J. Mol. Biol.*, **285**, 675–687.
22. Okamura,H., Hanaoka,S., Nagadoi,A., Makino,K. and Nishimura,Y. (2000) Structural comparison of the PhoB and OmpR DNA-binding/transactivation domains and the arrangement of PhoB molecules on the phosphate box. *J. Mol. Biol.*, **295**, 1225–1236.
23. Blanco,A.G., Sola,M., Gomis-Ruth,F.X. and Coll,M. (2002) Tandem DNA recognition by PhoB, a two-component signal transduction transcriptional activator. *Structure*, **10**, 701–713.
24. Bardin,S., Dan,S., Osteras,M. and Finan,T.M. (1996) A Phosphate transport system is required for symbiotic nitrogen fixation by *Rhizobium meliloti*. *J. Bacteriol.*, **178**, 4540–4547.
25. Bardin,S.D. and Finan,T.M. (1998) Regulation of Phosphate assimilation in *Rhizobium (Sinorhizobium) meliloti*. *Genetics*, **148**, 1689–1700.
26. Voegele,R.T., Bardin,S. and Finan,T.M. (1997) Characterization of the *Rhizobium (Sinorhizobium) meliloti* high- and low-affinity Phosphate uptake systems. *J. Bacteriol.*, **179**, 7226–7232.
27. Yuan,Z.C., Zaheer,R. and Finan,T.M. (2006) Regulation and properties of PstSCAB, a high-affinity, high-velocity, phosphate transport system of *Sinorhizobium meliloti*. *J. Bacteriol.*, **188**, 1089–1102.
28. Bardin,S.D., Voegele,R.T. and Finan,T.M. (1998) Phosphate assimilation in *Rhizobium (Sinorhizobium) meliloti*: identification of a *pit*-like gene. *J. Bacteriol.*, **180**, 4219–4226.
29. Parker,G.F., Higgins,T.P., Hawkes,T. and Robson,R.L. (1999) *Rhizobium (Sinorhizobium) meliloti phn* genes: characterization and identification of their protein products. *J. Bacteriol.*, **181**, 389–395.
30. Ruberg,S., Puhler,A. and Becker,A. (1999) Biosynthesis of the exopolysaccharide galactoglucan in *Sinorhizobium meliloti* is subject to a complex control by the Phosphate-dependent regulator PhoB and the proteins ExpG and MucR. *Microbiology*, **145**, 603–611.
31. Krol,E. and Becker,A. (2004) Global transcriptional analysis of the Phosphate starvation response in *Sinorhizobium meliloti* strains 1021 and 2011. *Mol. Gene Genomics*, **272**, 1–17.
32. McLean,P.A., Liu,C.M., Sookdeo,C.C. and Cannon,F.C. (1992) Characterization of a gene cluster involved in utilization of glyphosate and other Phosphonates in *Rhizobium meliloti*. GenBank accession no. M96263.
33. Tommassen,J., Koster,M. and Overduin,P. (1987) Molecular analysis of the promoter region of the *Escherichia coli* K-12 *phoE* gene. Identification of an element, upstream from the promoter, required for efficient expression of PhoE protein. *J. Mol. Biol.*, **198**, 633–641.
34. Kasahara,M., Makino,K., Amemura,M., Nakata,A. and Shinagawa,H. (1991) Dual regulation of the *ugp* operon by phosphate and carbon starvation at two interspaced promoters. *J. Bacteriol.*, **173**, 549–558.
35. Kim,S.K., Makino,K., Amemura,M., Shinagawa,H. and Nakata,A. (1993) Molecular analysis of the *PhoH* gene, belonging to the Phosphate regulon in *Escherichia coli*. *J. Bacteriol.*, **175**, 1316–1324.
36. Schneider,T.D., Stormo,G.D., Gold,L. and Ehrenfeucht,A. (1986) Information content of binding sites on nucleotide sequences. *J. Mol. Biol.*, **188**, 415–431.
37. Reeve,W.G., Tiwari,R.P., Worsley,P.S., Dilworth,M.J., Glenn,A.R. and Howieson,J.G. (1999) Constructs for insertional mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria. *Microbiology*, **145**, 1307–1316.
38. Oke,V. and Long,SR. (1999) Bacterial genes induced within the nodule during the *Rhizobium*-legume symbiosis. *Mol. Microbiol.*, **32**, 837–849.
39. Charles,T.C. and Finan,T.M. (1991) Analysis of a 1600-kilobase *Rhizobium meliloti* megaplasmid using defined deletions generated *in vivo*. *Genetics*, **127**, 5–20.
40. Finan,T.M., Hartweig,E., LeMieux,K., Bergman,K., Walker,G.C. and Signer,E.R. (1984) General transduction in *Rhizobium meliloti*. *J. Bacteriol*, **159**, 120–124.
41. Watson,R.J., Heys,R., Martin,T. and Savard,M. (2001) *Sinorhizobium meliloti* cells require biotin and either cobalt or methionine for growth. *Appl. Environ. Microbiol.*, **67**, 3767–3770.
42. Siehnel,R.J., Worobec,E.A. and Hancock,R.E. (1988) Regulation of components of the *Pseudomonas aeruginosa* phosphate-starvation-inducible regulon in *Escherichia coli*. *Mol. Microbiol.*, **2**, 347–352.
43. Nikata,T., Sakai,Y., Shibat,K., Kato,J., Kuroda,A. and Ohtake,H. (1996) Molecular analysis of the phosphate-specific transport (*pst*) operon of *Pseudomonas aeruginosa*. *Mol. Gen. Genet.*, **250**, 692–698.
44. Wu,H., Kato,J., Kuroda,A., Ikeda,T., Takiguchi,N. and Ohtake,H. (2000) Identification and characterization of two chemotactic transducers for inorganic phosphate in *Pseudomonas aeruginosa*. *J. Bacteriol.*, **182**, 3400–3404.
45. Rosenberg,H., Gerdes,R.G. and Chegwidden,K. (1977) Two systems for the uptake of phosphate in *Escherichia coli*. *J. Bacteriol.*, **131**, 505–511.
46. Yuan,Z.C., Zaheer,R. and Finan,T.M. (2005) Phosphate limitation induces catalase expression in *Sinorhizobium meliloti, Pseudomonas aeruginosa* and *Agrobacterium tumefaciens*. *Mol. Microbiol.*, **58**, 877–894.
47. Chao,T.C., Buhrmester,J., Hansmeier,N., Puhler,A. and Weidner,S. (2005) Role of the regulatory gene *rirA* in the transcriptional response of *Sinorhizobium meliloti* to iron limitation. *Appl. Environ. Microbiol.*, **71**, 5969–5982.
48. Geiger,O., Rohrs,V., Weissenmayer,B., Finan,T.M. and Thomas-Oates,J.E. (1999) The regulator gene *phoB* mediates Phosphate stress-controlled synthesis of the membrane lipid diacylglyceryl-N,N,N-trimethylhomoserine in *Rhizobium (Sinorhizobium) meliloti*. *Mol. Microbiol.*, **32**, 63–73.
49. Lopez-Lara,I.M., Sohlenkamp,C. and Geiger,O. (2003) Membrane lipids in plant-associated bacteria: their biosyntheses and possible functions. *Mol. Plant Microbe Interact.*, **16**, 567–579.
50. Klug,R.M. and Benning,C. (2001) Two enzymes of diacylglyceryl-O-4′-(N,N,N,-trimethyl) homoserine biosynthesis are encoded by *btaA* and *btaB* in the purple bacterium *Rhodobacter sphaeroides*. *Proc. Natl Acad. Sci. USA*, **98**, 5910–5915.
51. Lopez-Lara,I. M., Gao,J.L., Soto,M.J., Solares-Perez,A., Weissenmayer,B., Sohlenkamp,C., Verroios,G.P., Thomas-Oates,J. and Geiger,O. (2005) Phosphorus-free membrane lipids of *Sinorhizobium meliloti* are not required for the symbiosis with alfalfa but contribute to increased cell yields under phosphorus-limiting conditions of growth. *Mol. Plant Microbe. Interact.*, **18**, 973–982.
52. Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S., Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K. *et al.* (2000) Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res.*, **7**, 331–338.
53. Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E., Eisen,J.A., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R. *et al.* (2001) Complete genome sequence of *Caulobacter crescentus*. *Proc. Natl Acad. Sci. USA*, **98**, 4136–4141.
54. Kornberg,A., Rao,N.N. and Ault-Riche,D. (1999) Inorganic polyphosphate: a molecule of many functions. *Annu. Rev. Biochem.*, **68**, 89–125.