J Mol Evol (2005) 60:462–474
DOI: 10.1007/s00239-004-0212-7

JOURNAL OF MOLECULAR EVOLUTION

Comparison of the PhoPQ Regulon in *Escherichia coli* and *Salmonella typhimurium*

Pieter Monsieurs,[1] Sigrid De Keersmaecker,[2] William W. Navarre,[3] Martin W. Bader,[4] Frank De Smet,[1] Michael McClelland,[5] Ferric C. Fang,[3,4] Bart De Moor,[1] Jos Vanderleyden,[2] Kathleen Marchal[2,1]

[1] ESAT-SCD, K.U. Leuven, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium
[2] Centre of Microbial and Plant Genetics, K.U. Leuven, Kasteelpark Arenberg 20, 3001 Leuven-Heverlee, Belgium
[3] Department of Laboratory Medicine, University of Washington, Box 357110, HSB K-455, 1959 NE Pacific Street, Seattle, WA 98195–7710, USA
[4] Department of Microbiology, University of Washington, Box 357242, HSB K-155, 1959 NE Pacific Street, Seattle, WA 98195–7720, USA
[5] Sidney Kimmel Cancer Center, San Diego, CA 92121, USA

Received: 9 July 2004 / Accepted: 20 October 2004 [*Reviewing Editor:* Dr. Yves Van de Peer]

Abstract. The PhoPQ two-component system acts as a transcriptional regulator that responds to Mg^{2+} starvation both in *Escherichia coli* and *Salmonella typhimurium* (Garcia et al. 1996; Kato et al. 1999). By monitoring the availability of extracellular Mg^{2+}, this two-component system allows *S. typhimurium* to sense the transition from an extracellular environment to a subcellular location. Concomitantly with this transition, a set of virulence factors essential for survival in the intracellular environment is activated by the PhoPQ system (Groisman et al. 1989; Miller et al. 1989). Compared to nonpathogenic strains, such as *E. coli* K12, the PhoPQ regulon in pathogens must contain target genes specifically contributing to the virulence phenotype. To verify this hypothesis, we compared the composition of the PhoPQ regulon between *E. coli* and *S. typhimurium* using a combination of expression experiments and motif data. PhoPQ-dependent genes in both organisms were identified from PhoPQ-related microarray experiments. To distinguish between direct and indirect targets, we searched for the presence of the regulatory motif in the promoter region of the identified PhoPQ-dependent genes. This allowed us to reconstruct the direct PhoPQ-dependent regulons in *E. coli* K12 and *S. typhimurium* LT2. Comparison of both regulons revealed a very limited overlap of PhoPQ-dependent genes between both organisms. These results suggest that the PhoPQ system has acquired a specialized function during evolution in each of these closely related species that allows adaptation to the specificities of their lifestyles (e.g., pathogenesis in *S. typhimurium*).

Key words: PhoPQ regulon — *Escherichia coli* — *Salmonella typhimirium* — Genome evolution — Comparative genomics

Correspondence to: Kathleen Marchal; *email:* Kathleen.Marchal@agr.kuleuven.ac.be

Introduction

The PhoPQ regulatory system governs the adaptation to Mg^{2+} limiting conditions in *Salmonella typhimurium* and *Escherichia coli*. In this two-component system, the PhoQ protein acts as a sensor for extracytoplasmatic Mg^{2+} and Ca^{2+} (Vescovi et al. 1997) that controls the activity of the response regulator PhoP. This two-component system has also been shown to determine the virulence characteristics of *S. typhimurium* and other Gram-negative species (Groisman 2001). In *S. typhimurium*, the sensing of extracellular Mg^{2+} permits the pathogen to determine its subcellular location, i.e., inside macrophages and to activate the virulence factors essential for survival. Mutations in the phoPQ operon of *S. typhimurium*

result in an attenuated virulence phenotype i.e., an increased sensitivity to cationic antimicrobial peptides (Fields et al. 1989; Groisman et al. 1992; Guo et al. 1997; Gunn et al. 1998; Guina et al. 2000; Bader et al. 2003) and acid pH (Foster and Hall 1990; Bearson et al. 1998), a decreased resistance to bile salts (van Velkinburgh and Gunn 1999; Prouty and Gunn 2000), a deficiency in epithelial cell invasion, and the inability to survive within macrophages (Fields et al. 1986; Miller et al. 1989; Darwin and Miller 1999). However, although the PhoPQ system is a major determinant of the virulence characteristics of *S. typhimurium*, it does not confer a similar phenotype in a nonpathogenic strain of *E. coli*. In *E. coli* K12, the PhoPQ two-component system is, besides its physiological role in response to Mg^{2+} starvation, important for resistance to antimicrobial peptides (Groisman et al. 1992; Groisman 2001). The amino acid sequences of PhoP and PhoQ of *E. coli* are 93 and 86% identical, respectively, to those of *S. typhimurium* (Kasahara et al. 1992). This implies that the PhoPQ systems in *E. coli* and *S. typhimurium* are most probably functional counterparts of each other. A short consensus site (T/G)GTTTA, occurring as an interrupted dyad, has been suggested to be the binding site for the PhoP regulatory protein in *Salmonella* spp. (Soncini et al. 1995) and *Escherichia coli* (Kato et al. 1999). The biological relevance of this motif was experimentally verified by Yamamoto et al. (2002) and Lejona et al. (2003) in *E. coli* and *S. typhimurium*, respectively.

In this study, we estimated the size of the direct PhoPQ-dependent regulons in both species by combining the evidence gained from microarray data and motif information. Comparing the gene composition of the PhoPQ regulon revealed a very small overlap in the genes that were PhoPQ regulated in both species, indicating that the difference in virulence phenotype between the pathogenic and the nonpathogenic bacteria might be attributed to a group of target genes that are specifically PhoP regulated in the pathogenic *S. typhimurium* strain but not in *E. coli* K12.

Materials and Methods

First, PhoP-dependent operons were identified in *S. typhimurium* based on the microarray dataset of this study and in *E. coli* based on a previously published dataset (Minagawa et al. 2003). Subsequently, promoter regions of these PhoPQ-dependent operons were screened for the presence of the PhoP regulatory motif. This motif information was used to distinguish in each of the respective organisms the direct from the indirect regulon. Eventually, overlap between the regulons in both organisms was identified.

Identification of PhoPQ-Dependent Genes Based on Expression Data

To identify genes dependent on PhoPQ in *E. coli*, we used the results described by Minagawa et al. (2003). They used cDNA microarray experiments to identify target genes of the Mg^{2+} stimulon that responded to the availability of external Mg^{2+} in a PhoPQ-dependent manner. In their study, wild-type and *phoP*- and *phoQ*-defective *E. coli* strains were grown in the presence and absence of Mg^{2+}, respectively. For the identification of the PhoPQ regulon in *S. typhimurium*, we set up a microarray experiment (data available at the supplementary information) in which RNA was isolated from strains of *Salmonella enterica* serovar Typhimurium ATCC 14028 harboring either a null mutation (*phoP::Tn10dCm*) or a constitutive mutation (*phoQ24*) in the genes encoding the PhoPQ two-component system (Miller and Mekalanos 1990). Strains were grown to mid-log phase in M9 minimal medium prior to harvesting RNA and the array experiments were performed as described by Bader et al. (2003).

Data were statistically analyzed by combining six replicates for each experiment using the maximum-likelihood analysis of Ideker et al. (2000, 2001). This method calculates for each gene a generalized likelihood ratio test statistic (λ value) and the ratio of the mean intensities for the two conditions (μ-ratio). The λ values were converted into p-values using a χ^2 cumulative probability distribution with one degree of freedom (Ideker et al. 2000). These p-values were used to estimate the number of actually differentially expressed genes (Fig. 1A) and subsequently to plot the false discovery rate (FDR) and the number of true positives for each critical value of λ (Figs. 1A and 1B). Detailed description on how to interpret these plots can be obtained from Storey and Tibshirani (2003) and De Smet et al. (2004). In order to minimize the type 1 error (number of false positives), we chose a high critical threshold of λ (λ cutoff of 40) resulting in 324 significantly differentially expressed genes. Under the assumption that genes directly regulated by PhoPQ will be most severely affected by the PhoP mutation, we selected from the 324 most significantly differentially expressed genes (using a λ cutoff of 40) those for which the μ-ratio was less than −0.30 (genes upregulated by PhoP). This additional selection step reduced this number of selected genes to a subset of the 214 most influenced genes. This number approximated the number of *E. coli* PhoPQ upregulated genes selected by Minagawa et al. (2003).

Sequence-Based Analysis

For sequence analysis, genome sequences were obtained from GenBank (NC_003197 for *S. typhimurium* LT2, NC_003277 for *S. typhimurium* LT2 plasmid, and NC_000913 for *E. coli* K12) (Benson et al. 2003). Intergenic regions used in this study were extracted from the genome sequences using the modules implemented in INCLUSive (Thijs et al. 2001) to automatically parse GenBank files. Here, we define intergenic regions as the noncoding sequence between two coding sequences. No overlap was allowed with the coding sequences.

Identification of Orthologs. Clusters of orthologs were identified using TribeMCL (Enright et al. 2002). The obtained clusters contain proteins that are close homologs (either orthologs or paralogs) of each other. The pairwise BLAST scores obtained by mutually aligning the whole-genome sequences using blastp (Altschul et al. 1997) were used as input of TribeMCL. Stringent criteria were applied to retain only closely related orthologs and paralogs (cutoff of the BLAST hit was an e-value of $1e^{-80}$). When the BLAST hit of a protein against itself resulted in an expectation value higher than $1e^{-80}$, we used less stringent criteria (a cutoff of $1e^{-20}$). Because the analysis of the PhoPQ regulon is performed at an operon level, in the context of this study we defined an operon to be unique if, for all genes of an operon in one organism (e.g., *E. coli*), no orthologs are found in the other organism (e.g., *S. typhimurium*).

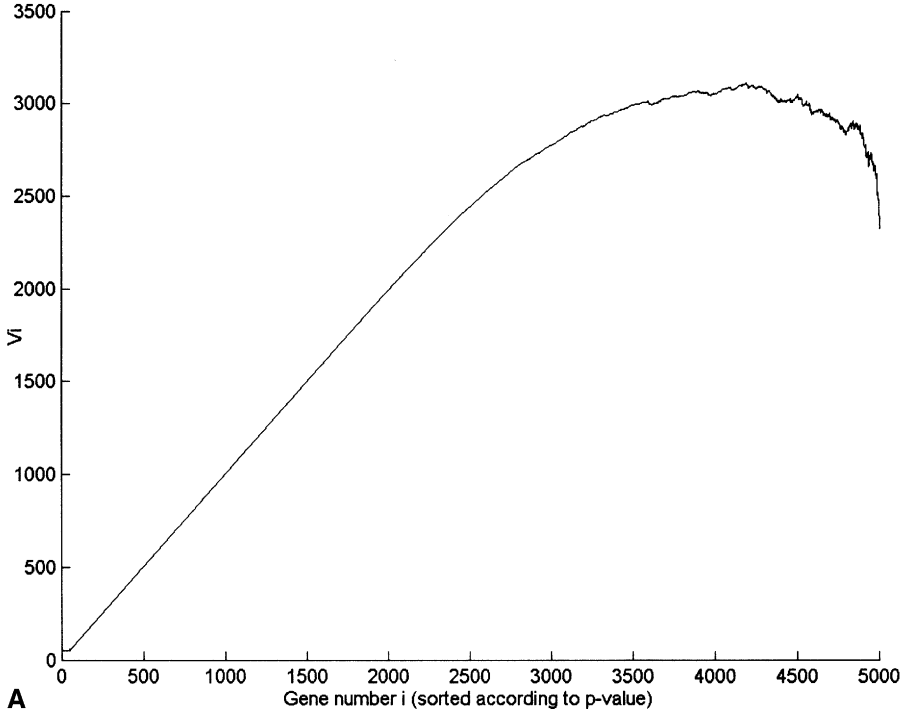


Fig. 1A. Plot of V_i versus the gene number (sorted according to p-value).Where n is the total number of genes in the dataset, p is the p-value of gene i, i is the rank order of a gene after sorting all genes according to their p-value, and V_i reaches a constant level at 2855 genes, which is an estimate for the number of actually differentially expressed genes.

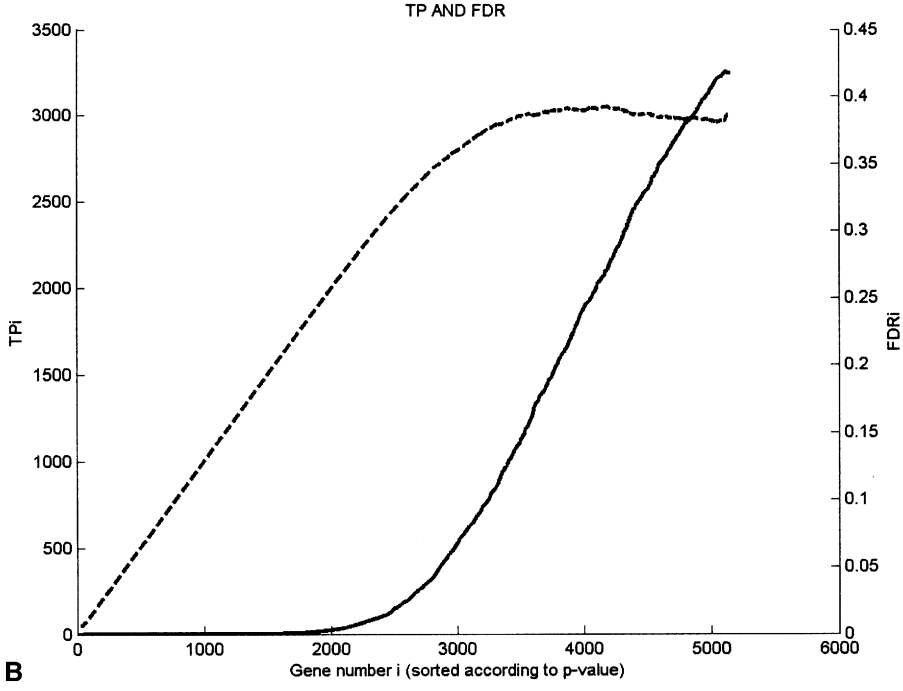


Fig. 1B. Number of true positives (TP, dashed line)/FDR (solid line) versus the gene number (genes are sorted according to their p-value).

Operon Prediction. Note that for our analysis it was of importance not to detect highly reliable operons (high selectivity, low sensitivity), but to unveil anything that could possibly be an operon (high sensitivity, low selectivity). Therefore, compared to studies focusing on operon prediction, we used deliberately non-stringent criteria for our operon prediction: the intergenic distance was used as the main criterion for defining an operon (Salgado et al. 2000; Moreno-Hagelsieb and Collado-Vides 2002). When two genes were located on the same strand and the distance of the intergenic sequence between these two genes was smaller than a predefined cutoff value, these two genes were considered residing in the same operon. For both *E. coli* and *S. typhimurium* we used a cutoff value of 40 nucleotides to predict operons.

These cutoff values were determined as follows: we first predicted operons in *E. coli* using different cutoff values. Each of these predicted sets was compared to the operon prediction of RegulonDB that was used as a benchmark set (Salgado et al. 2004). An operon in RegulonDB was considered to be identical to our own

predicted operon if both operons contained exactly the same genes. A maximal match of the predicted operons in *E. coli* with the RegulonDB database was obtained using the cutoff value of 40 nucleotides, mentioned above. Since for *S. typhimurium* no such benchmark dataset existed, we used the cutoff value on the *S. typhimurium* intergenic distance that maximized the match between the predicted set in *E. coli* and that in *S. typhimurium*. A match in this context means that an operon in *E. coli* and *S. typhimurium* contains orthologous genes only. For *E. coli* the average number of genes in an operon was predicted to be 2.88, while for *S. typhimurium* an average of 2.90 genes per operon was found.

Motif Detection. We used MotifSampler (Thijs et al. 2002) to construct the PhoP motif model (i.e., a probabilistic representation of the DNA pattern). The MotifSampler is a motif detection algorithm based on Gibbs sampling that allows retrieving statistically overrepresented patterns in the promoter regions of coregulated genes (Thijs et al. 2001; Marchal et al. 2003). Based on a training set containing promoter regions of known PhoPQ-regulated genes (*phoP*, *mgtA*, *pmrD, pdgL*, and *slyB* in *S. typhimurium* [Yamamoto et al. 2002; Lejona et al. 2003] and *yrbL, ybcU/vboR*, and *rstA* in *E. coli* [Minagawa et al. 2003]), the motif detected with the highest information content had a length of six nucleotides and a consensus site (T/G)GTTTA. This corresponds to one half site of the previously suggested dyad PhoP motif (Soncini et al. 1995; Yamamoto et al. 2002). The motif model corresponding to this motif was used for the genomewide motif screening of all intergenic sequences of *E. coli* and *S. typhimurium*, respectively, using MotifLocator (Thijs 2003; Marchal et al. 2004). This algorithm uses the motif model to calculate a score for each window (with length similar to the length of the motif model) in the intergenic sequences. The threshold score was set at 0.75 (nonstringent screening criterion). This threshold corresponds to the selection of approximately the 2% best scoring motif hits of all possible motif positions in the genome if the intergenic regions of the whole genome of *S. typhimurium* and *E. coli*, respectively, would be screened (*S. typhimurium*, 1.91%; and *E. coli*, 1.96%).

For the stringent screening criterion, we included extra restrictions on the PhoP motif model that were derived from previous publications (Yamamoto et al. 2002; Minagawa et al. 2003; Lejona et al. 2003), in which the PhoP motif was described as a direct repeat of (T/G)GTTTA. A conserved thymine in the first half site (at position 3) and two conserved thymines and one conserved adenine in the second half site (at positions 3, 4, and 6, respectively) were shown to be essential for the binding of the PhoP regulatory protein. Therefore, as an additional criterion we explicitly required that these four positions were conserved inside and around the motif instances retrieved from the first screening.

Calculating the Overlap in Regulon Composition

To identify the direct regulon, we checked whether the intergenic regions of the first genes of operons that were selected as differentially expressed in either one (or both) of the organisms contained a PhoP motif. To prove unequivocally the presence of a restricted overlap in regulon composition of the PhoPQ system between the two organisms compared (see Discussion), we used extremely nonstringent criteria to calculate this overlap: for each gene differentially expressed by PhoP in one organism, we first identified the operon to which it belonged (based on the operon prediction outlined above). If at least one of the orthologs of the genes belonging to that operon turned out to be differentially expressed in the other organism as well, the operon was considered differentially expressed in both organisms (i.e., identifying an operon as PhoPQ dependent in both organisms does not necessarily imply that all genes of the operon were identified as differentially expressed in both organisms).

Enrichment of Functional Classes

To identify which functional classes were enriched in the directly PhoPQ-dependent *E. coli* and *S. typhimurium* genes, the functional classifications of, respectively, EcoCyc (Karp et al. 2004) and the *S. typhi* Sanger database (Parkhill et al. 2001) were used. To use the annotation of the Sanger database of *S. typhi* strain C18 for *S. typhimurium*, a mapping between the *S. typhi* and the *S. typhimurium* gene names based on ortholog sequence information was made. Functional enrichment of PhoPQ-dependent targets was calculated using the hypergeometric distribution (Tavazoie et al. 1999), which assigns to each functional class a *p*-value. This *p*-value describes for each functional class the probability that in a random set of genes, the same number of genes of that specific functional class will be observed. In detail, the probability of observing at least k predicted PhoPQ-regulated genes from a functional category within the total number of predicted PhoPQ target genes (n) is given by

$$P = 1 - \sum_{i=0}^{k-1} \frac{\binom{f}{i}\binom{g-f}{n-i}}{\binom{g}{n}} \quad (1)$$

where f is the total number of genes in the functional category in the EcoCyc database or the *S. typhi* Sanger database and g is the total number of genes within the genome of *E. coli* and *S. typhimurium*, respectively.

Supplementary Information

Supplementary material is available at ftp://ftp.esat.kuleuven.ac.be/pub/sista/marchal/Supplementary_informationPhoPQ_2004/SuppWebsite.html.

Results

Identification of the PhoP-Dependent Regulon Based on Microarray Analysis

To identify genes dependent on PhoPQ in *E. coli*, we used the results described by Minagawa et al. (2003). They identified 219 genes as being upregulated by PhoP.

To identify the regulon regulated by PhoPQ in *S. typhimurium* we compared expression in a knockout and constitutive mutant of the genes encoding the PhoPQ system. Based on the analysis summarized in Fig. 1A the expression of approximately 2855 genes was found to be affected to some extent by the constitutive mutation. Based on selection criteria outlined in Materials and Methods, we selected, from the 324 most significantly differentially expressed genes, the 214 most upregulated genes. This number approximated the number of *E. coli* PhoPQ-upregulated genes selected by Minagawa et al. (2003). Conclusively, for both datasets we obtained a subset of genes most severely upregulated by PhoPQ. The quality of the microarray results and the biological relevance of the used cutoff values were confirmed by the presence of experimentally verified PhoPQ regulated genes in the selected subset of genes.

Based on our analysis, the 219 genes differentially PhoP regulated in *E. coli* were organized in 193 operons (on a total of 2848 predicted operons, of which 2027 are singletons). Similarly, the 214 most affected *S. typhimurium* genes were located in 189 operons (on a total of 3026 predicted operons of which 2195 are singletons).

Motif Screening

To identify target operons directly regulated by the PhoPQ regulatory system, intergenic regions of the first genes of the operons that were identified as differentially expressed in either one or both organisms were screened with the PhoP motif model. Initially the PhoP box was postulated to be a direct repeat of the hexanucleotide (T/G)GTTTA separated by a spacer of five nucleotides. Recent experimental evidence allowed refinement of the specific sequence requirements of the PhoP box (Yamamoto et al. 2002; Minagawa et al. 2003; Lejona et al. 2003). These studies confirmed that the PhoP regulatory protein is able to bind to a promoter region that does not display an intact dyad motif. A conserved thymine in the first half site (at position 3) and two conserved thymines together with one conserved adenine in the second half site (at positions 3, 4, and 6, respectively) were sufficient for detection of the DNA binding of the PhoP protein in *in vitro* DNA footprinting analysis.

However, as mentioned in these experimental studies, it cannot be excluded that alterations at some of these positions just reduce the PhoPQ-dependent regulation to a level that cannot be observed by *in vitro* footprinting analysis (Lejona et al. 2003). Therefore, in order not to bias our initial genomewide screening toward previously made observations and to guarantee a maximal sensitivity, we first screened the selected promoter regions with a motif model that describes only one half site of the PhoP box (high-sensitive, low-specific screening). In addition, we also performed a second more stringent screening taking into account also the specific sequence requirements described above (low-sensitive, high-specific screening). The first screening resulted in 130 and 147 putative target operons in *E. coli* and *S. typhimurium*, respectively. Taking into account the more stringent criterion, 42 and 34 operons potentially directly dependent on PhoP were retained in *S. typhimurium* and *E. coli*, respectively (see Fig. 2A and Fig. 2B).

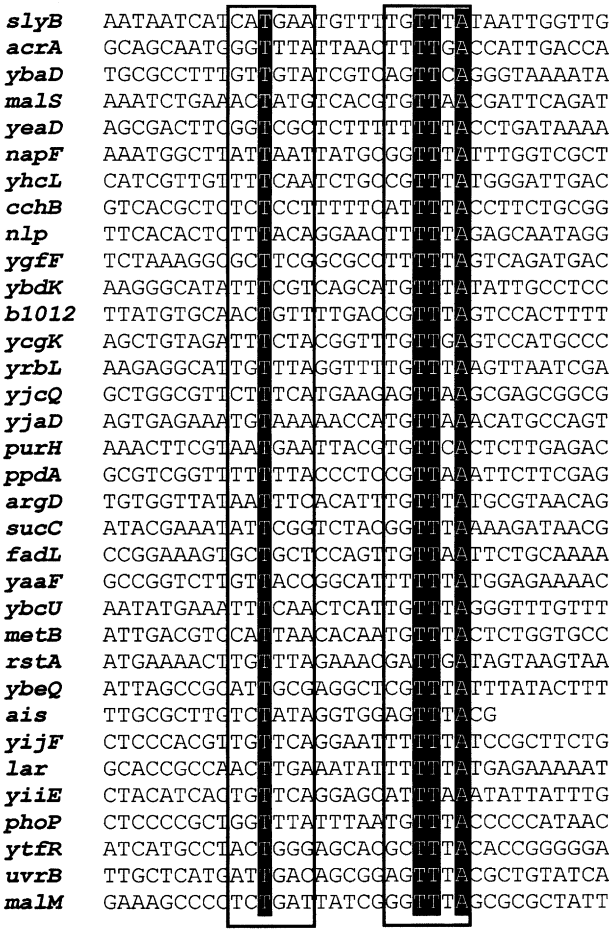
Combination of Motif Screening Results and Microarray Data

In this section we compare the overlap of the PhoPQ regulon between *E. coli* and *S. typhimurium,* i.e., between the 193 operons of *E. coli* and the 189 operons of *S. typhimurium* that were identified as upregulated by PhoPQ in each of the respective organisms. The directly regulated operons were distinguished from the indirectly regulated ones by the presence of a PhoP motif in the region upstream of a gene. Results are based on the stringent screening criteria. Detailed information can be found at the supplementary Web site. A schematic representation is shown in Fig. 3.

Corresponding Genes in Both Datasets. When comparing both datasets, 13 operons were differentially expressed in both the *S. typhimurium* and *E. coli* (Fig. 3A). Only 2 of these 13 orthologous operons contained a PhoP box in their upstream region in both *S. typhimurium* gene and *E. coli* (i.e., *phoPQ* and *slyB*). In both these genes, a PhoP motif was also conserved in the evolutionary-related bacteria *Shigella flexneri* and *Yersinia pestis*, offering an extra validation for the biological relevance of the identified regulatory motif. In one gene (*pagP*), a PhoP box was found in *S. typhimurium* but was missing in the promoter region of the corresponding ortholog of *E. coli* (*crcA*). These results suggest that there is a direct regulation of *pagP* in *S. typhimurium*, while this regulation is indirect in *crcA* in *E. coli*. Alternatively, this motif might be a false-positive result. In the remaining 10 operons, no PhoP box was present either in *E. coli* or in *S. typhimurium*, indicating that these genes are probably indirectly regulated by the PhoPQ two-component system. A detailed list of the operons that are differentially expressed in both organisms can be found in Table 1.

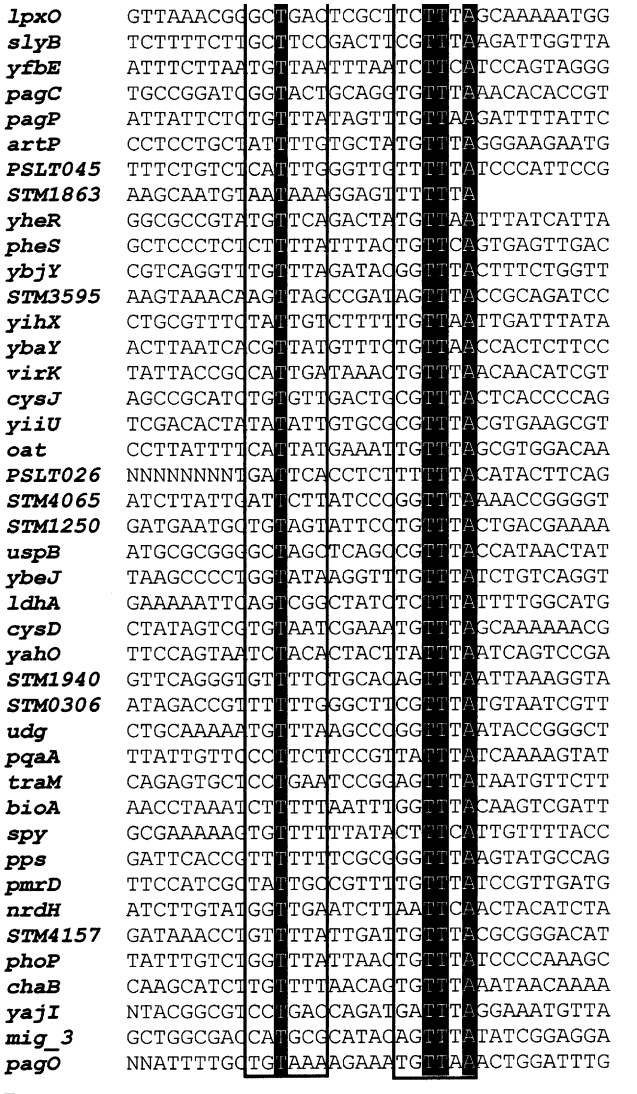
E. coli *Microarray Data.* Besides the 13 operons that were differentially expressed in both organisms according to the corresponding microarray data, remarkable differences were found for other operons (Fig. 3B). Forty-five operons that were differentially expressed in *E. coli* do not have an orthologous operon in *S. typhimurium* or in evolutionary-related bacteria like *Yersinia pestis* and *Shigella flexneri*. Eleven of these operons displayed a PhoP box in their promoter region, while the remaining 34 *E. coli* operons were most likely indirectly regulated by the PhoPQ system (no PhoP box found).

On the other hand, 135 *E. coli* operons that were differentially expressed in the microarray of Minagawa et al. (2003) and that had an orthologous operon in *S. typhimurium* were not differentially expressed according to our data in *S. typhimurium*. For 17 of these 135 orthologous operons, differential expression and direct regulation (i.e., a PhoP box found) was confined to *E. coli* only. Ninety-five of these 135 Mg^{2+} repressed *E. coli* operons appeared to be indirectly regulated by the PhoPQ two-component system (no PhoP box was detected upstream these



A

Fig. 2A. Alignment of the detected PhoP motifs in *E. coli*. Conserved positions are highlighted.



B

Fig. 2B. Alignment of the detected PhoP motifs in *S. typhimurium*. Conserved positions are highlighted.

operons). Besides the previous results, some inconsistent observations were made. Nineteen of the PhoPQ-dependent differentially expressed *E. coli* operons did not contain a motif, while in the corresponding non-differentially expressed *S. typhimurium* orthologs, a PhoP box was present. For a total of four operons, in both organisms a PhoP box was found but differential expression was observed in *E. coli* only. The latter two classes might be due to false-positive motif hits, measurement errors in the microarray data or incomplete sampling of relevant PhoPQ-dependent conditions. Therefore, these results should be interpreted with caution.

S. typhimurium *Microarray Data*. Performing a similar analysis for *S. typhimurium* (Fig. 3C), 84 PhoPQ-dependent *S. typhimurium* operons do not have an ortholog in *E. coli* or in evolutionary-related species. In 21 of these 84 unique operons, we found evidence for direct PhoPQ dependency (i.e., a PhoP box is found in the promoter region of the first gene of the operon), while the remaining 63 operons were indirectly PhoPQ dependent (i.e., no PhoP box present in the promoter region of the gene). Ninety-two of the significantly differentially expressed *S. typhimurium* operons contained an orthologous operon in *E. coli*. However, these operons in *E. coli* seemed not to be differentially expressed in a PhoPQ-dependent way. In 13 of these 92 *S. typhimurium* operons, evidence for direct PhoPQ-dependent regulation was found. On the other hand, 64 of the 92 operons did not display a PhoP motif in their promoter region pointing toward indirect regulation by the PhoPQ system in *S. typhimurium* operons. Also in this dataset, a limited number of inconsistencies was observed for which no clear conclusions could be drawn. For 9 of the 92 PhoPQ-dependent differentially expressed *S. typhimurium* operons that did not contain a PhoP motif, a PhoP box was present in the corresponding non-differentially regulated *E. coli*

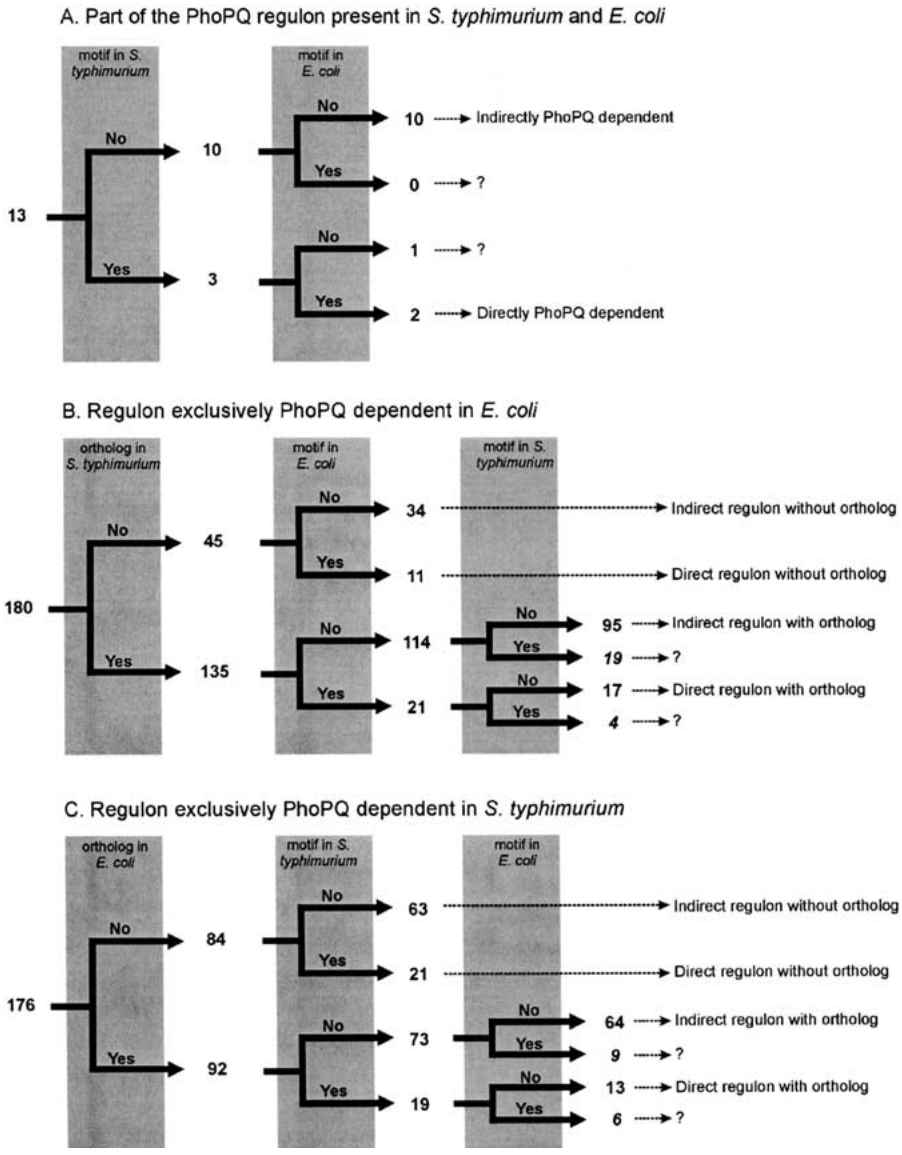


Fig. 3. **A** Overview of the PhoPQ-dependent genes differentially expressed in both *S. typhimurium* and *E. coli*. Each branch of the tree corresponds to a specific subcategory. The number of operons for the different subcategories is displayed. Subcategories for which the results need to be interpreted with caution are indicated by a ?. **B** Overview of the PhoPQ-dependent genes differentially expressed in *E. coli* but not in *S. typhimurium*. Legend as for A. **C**. Overview of the PhoPQ-dependent genes differentially expressed in *S. typhimurium* but not in *E. coli*. Legend as for B.

orthologs. Another six operons contained a PhoP box in both organisms, although differential expression only was observed in *S. typhimurium*.

Functional Classes

Based on the functional classification schemes of the EcoCyc and the *S. typhi* Sanger database, we retrieved for each PhoPQ regulated gene in *E. coli* and *S. typhimurium* its corresponding functional class. We subsequently determined which functional classes were significantly enriched in the sets of directly/indirectly PhoPQ-dependent genes in both *E. coli* and *S. typhimurium*. It should be mentioned that many of the detected PhoPQ-dependent targets (especially the unique genes) are not yet annotated in these databases, resulting most probably in an underestimation of the degree of overrepresentation of certain functional classes.

For both the stringent and the nonstringent estimation of the direct PhoPQ-dependent regulon, we calculated the enrichment of functional categories in the selected gene sets. The enrichment was most pronounced using the gene sets resulting from the stringent criteria (lower *p*-values). Moreover, the functional classes that were enriched in these gene sets better corresponded to the known functionalities of PhoPQ-dependent genes and the average size of the stringent direct PhoPQ-dependent regulon approximated the previously predicted direct PhoPQ regulon size of 40 loci (Soncini et al. 1996; Soncini and Groisman 1996). Therefore, the most stringent criteria seemed to be the best approximation of the true regulon composition and these were used for further characterization of this regulon.

In Table 2, the overrepresented functional classes in both the directly and the indirectly PhoPQ-dependent regulons are shown for *S. typhimurium*.

Table 1. List of PhoPQ-dependent operons that are differentially expressed both in *E. coli* and in *S. typhimurium*

E. Coli		*S. typhimurium*		
Operon	Motif	Operon	Motif	Function
galETK	No	*galETK*	No	Galactose metabolism
tktAB	No	*talA–tktB*	No	Transaldolase/transketolase
gcvTH	No	*gcvTH*	No	Glycine metabolism
ompT	No	*pgtE*	No	Outer membrane protein
atpFHA	No	*atpFHA*	No	Membrane bound ATP synthase
malFG	No	*malFG*	No	Maltose transport
dacC	No	*dacC*	No	D-Alanyl-D-alanine carboxypeptidase
yraNOP	No	*yraNOP*	No	Endonuclease–isomerase–periplasmatic protein
bioBFCD	No	*bioBFCD*	No	Biotine synthesis
cmtBA	No	*STM2344*	No	Transport of small molecules
crcA	No	*pagP*	Yes	Lipid A modification
slyB	Yes	*slyB*	Yes	Outer membrane protein
phoPO	Yes	*phoPQ*	Yes	Regulatory protein—magnesium sensitive

The categories of "small molecule metabolism" (subcategories "degradation" [1.A] and "energy metabolism" [1.B]), "protein translation and modification" (3.A.8), "transport of anions and carbohydrates, organic acids and alcohols" (4.A.3 and 4.A.5), and "adaptations and atypical conditions (5.F)" are mainly overrepresented in the indirect regulon. Overrepresentation of the classes "aminoacyl tRNA synthetases" (3.A.5 containing *pheS, pheT, rbn*) and "transport of amino acids and amines" (4.A.1; containing *artP, artI, nrdF, STM4157, artM, PSLT043, ybj*) and the class "drug/analogue sensitivity" (5.D; to which belong *pmrD, pqaA, pagP*) is confined to the direct regulon. Significantly enriched within the subset of both direct and indirect PhoPQ target genes is the category for "central intermediary metabolism" (1.C), e.g., the directly regulated genes of sulfur metabolism (*cysC, cysI, cysJ, cysN, cysD*), the functional class involved in the "synthesis and modification of the cell envelope" (3.C.1), to which belong the directly PhoPQ regulated genes *slyB, pmrF, yfbG, STM2303, STM1940, spy, STM1864, lpxO, STM2302, ybjY, ybaY, STM4065, yahO*, and *pagO*, and the functional category responsible for "cation transport" (4.A.2), to which belong the directly regulated genes *chaB, artQ*, and *kefB*.

In *E. coli*, a significant part of both direct and indirect dependent PhoPQ genes is also enriched in functional categories involved in general metabolism (7; e.g., "central intermediary metabolism and carbon utilization"), in "transport" (9), and in functions related to "cell wall and membrane structure" (2). Besides "these previous classes," the indirect regulon is also enriched in the categories "cell processes" (1) and information transfer (5). Some genes, mainly those of direct regulon, are involved in "regulation" (8). Note, however, that the functional analysis of these *E. coli* genes is only partial because 30 of the potentially directly regulated PhoPQ genes are not assigned to any functional class in the EcoCyc database.

Detailed Analysis of Novel Direct PhoPQ Targets

Most of the known PhoPQ regulated genes with an experimentally verified PhoP box could be retrieved by our analysis (*phoP, slyB, pdgL/pcgL*, and *pmrD* in *S. typhimurium* [Lejona et al. 2003] and *phoP, slyB, yrbL, ybcU/vboR*, and *rstA* in *E. coli* [Minagawa et al. 2003]). The *mgtA* gene was missing in our analysis because it was inaccurately measured (low significance) on our arrays and was not spotted on the microarray of Minagawa et al. (2003). Retrieving the known PhoPQ targets illustrates the predictive power of our analysis and allows us to suggest the presence of new, promising, directly PhoPQ regulated targets in *S. typhimurium*.

To the functional category related to the "cell envelope" belong several directly PhoPQ-regulated genes with, as the most promising novel targets, *pagO, lpxO*, and *STM1940*. *pagO* encodes an integral membrane protein, previously described as PhoPQ dependent (Gunn et al. 1998) that is similar to a product of the *Yersinia* virulence plasmid. According to our analysis, *pagO* would be directly PhoPQ dependent. *lpxO* codes for a dioxygenase that plays a role in lipid synthesis (Gibbons et al. 2000). Belonging to the same functional class (subcategory "membranes and lipoproteins") is the novel potential direct PhoPQ target STM1940. This gene codes for a cell wall-associated hydrolase in *S. typhimurium*.

The set of directly PhoPQ-dependent genes we identified was also statistically enriched for the functional class related to "drug sensitivity." Three of the genes we identified from this category (*pmrD, pqaB, pagP*) are important for resistance to antimicrobial peptides. The *pmrD* gene is shown to be directly PhoPQ regulated (Kato et al. 2003). The *pagP*

Table 2. Overview of enriched functional classes for PhoPQ-dependent *S. typhimurium* genes

Number	Description	Indirect	Direct
1	Small molecule metabolism	**0.10**	0.60
1.A	Degradation	**0.16**	1.00
1.A.1	Carbon compounds	**0.18**	1.00
1.*B*	Energy metabolism	**0.10**	0.97
1.B.1	Glycolysis	0.23	1.00
1.B.5	Pentose phosphate pathway	**0.00**	1.00
1.B.5.b	Nonoxidative branch	**0.00**	1.00
1.B.7	Respiration	**0.99**	1.00
1.B.7.a	Aerobic	**0.86**	1.00
1.B.7.b	Anaerobic	**0.98**	1.00
1.B.8	Fermentation	**0.00**	0.22
1.C	Central intermediary metabolism	**0.01**	**0.00**
1.C.2	Gluconeogenesis	1.00	**0.06**
1.C.3	Sugar nucleotides	**0.02**	**0.17**
1.C.5	Sulfur metabolism	1.00	**0.00**
1.D	Amino acid biosynthesis	0.78	1.00
1.D.2	Aspartate family	0.73	1.00
1.D.3	Serine family	0.36	1.00
1.D.4	Aromatic amino acid family	0.70	1.00
1.D.6	Pyruvate family	0.65	1.00
1.F	Purines, pyrimidines, nucleosides, and nucleotides	1.00	0.31
1.F.5	Miscellaneous nucleoside/nucleotide reactions	1.00	**0.01**
1.G	Biosynthesis of cofactors, prosthetic groups, and carriers	0.61	0.60
1.G.1	Biotin	0.32	0.33
1.G.10	Thioredoxin	0.42	**0.16**
1.G.9	Riboflavin	0.22	1.00
2	Broad regulatory functions	0.23	0.75
3	Macromolecule metabolism	0.74	0.30
3.A	Synthesis and modification of macromolecules	0.89	0.51
3.A.2	Ribosomal protein synthesis and modification	0.80	1.00
3.A.3	Ribosome maturation and modification	0.36	1.00
3.A.5	Aminoacyl tRNA synthetases and their modification	1.00	**0.04**
3.A.7	DNA replication, restriction/modification, recombination, and repair	0.95	0.77
3.A.8	Protein translation and modification	**0.10**	1.00
3.A.9	RNA synthesis, RNA modification, and DNA transcription	1.00	0.32
3.B	Degradation of macromolecules	0.88	1.00
3.B.3	Proteins, peptides, and glycopeptides	0.46	1.00
3.C	Cell envelope	0.39	**0.19**
3.C.1	Membranes, lipoproteins, and porins	**0.20**	**0.08**
3.C.2	Surface polysaccharides, lipopolysaccharides, and antigens	0.39	0.38
3.C.4	Murein sacculus and peptidoglycan	0.33	1.00
4	Cell processes	0.63	0.69
4.A	Transport/binding proteins	0.31	**0.14**
4.A.1	Amino acids and amines	1.00	**0.00**
4.A.2	Cations	**0.16**	**0.01**
4.A.3	Carbohydrates, organic acids, and alcohols	**0.04**	1.00
4.A.5	Anions	**0.01**	1.00
4.A.6	Other	0.81	1.00
4.C	Cell division	0.75	1.00
4.D	Chemotaxis and mobility	0.53	1.00
4.G	Detoxification	0.29	1.00
4.H	Cell killing	0.22	1.00
4.I	Pathogenicity	0.93	0.97
5	Other	0.28	0.28
5.A	IS elements, phage-related functions and prophage	1.00	0.54
5.D	Drug/analogue sensitivity	1.00	**0.01**
5.F	Adaptions and atypical conditions	**0.03**	1.00
5.H	Conserved hypothetical protein	0.21	0.92
5.H.a	Hypothetical protein (conserved in *E. coli*)	0.72	0.97
5.H.b	Hypothetical protein (conserved in organism[s] other than *E. coli*)	**0.10**	0.74
5. I	Unknown	0.43	**0.10**

Note. Columns 1 and 2: number and description of the functional categories according to the *S. typhi* Sanger database. Column 3: *p*-values for the indirect PhoPQ regulon, Column 4: *p*-values for the direct PhoPQ regulon genes. *p*-values describe the statistical significance of the enrichment of a functional class within the subset of PhoPQ-regulated genes. The most significant fields (*p*-value below 0.2) are boldfaced.

Table 3. Overview of enriched functional classes for PhoPQ-dependent *E. coli* genes

Number	Description	Indirect	Direct
1	**Cell processes**	**0.01**	**0.13**
	Adaptations	0.92	1.00
	Cell division	0.24	1.00
	Protection	0.38	**0.18**
2	**Cell structure**	**0.01**	**0.04**
	Flagella	**0.01**	1.00
	Membrane	**0.05**	**0.03**
	Murein	0.50	1.00
	Pilus	0.25	**0.01**
	Ribosomes	**0.12**	1.00
4	**Extrachromosomal**	0.21	0.68
	Transposon related	1.00	1.00
5	**Information transfer**	**0.08**	0.71
	DNA related	**0.11**	**0.18**
	Protein related	**0.01**	0.50
	RNA related	1.00	0.98
6	**Location of products**	1.00	1.00
	Cytoplasm	0.54	0.74
	Extracellular	**0.15**	1.00
	Inner membrane	0.31	**0.09**
	outer membrane	0.65	0.53
	Periplasm	0.67	**0.01**
7	**Metabolism**	**0.01**	**0.01**
	Biosynthesis building blacks	1.00	1.00
	Flagellum	**0.01**	1.00
	Large molecule carriers	0.86	**0.01**
	Lipopolysaccharide	**0.01**	1.00
	Biosynthesis macromolecules	1.00	1.00
	Carbon Utilization	0.64	0.26
	Amines	**0.19**	**0.01**
	Amino acids	**0.15**	**0.01**
	Carbon compounds	0.43	0.80
	Fatty acids	0.47	0.30
	Central intermediary metabolism	**0.02**	0.46
	Energy metabolism	1.00	1.00
	Aerobic respiration	**0.01**	1.00
	Anaerobic respiration	0.01	0.01
	Fermentation	0.10	1.00
	Glycolysis	0.77	1.00
	TCA cycle	**0.19**	**0.02**
	Energy productions/transport	1.00	1.00
	Electron acceptors	**0.03**	**0.04**
	Electron carriers	**0.04**	**0.01**
	Electron donors	**0.00**	1.00
8	**Regulation**	0.24	**0.07**
	Genetic unit regulated	0.99	0.85
	type of regulation	0.62	**0.14**
	DNA structure level	0.60	1.00
	Posttranscriptional	0.63	0.28
	Unknown	**0.13**	1.00
9	**Transport**	**0.00**	**0.11**
	Group translocators	**0.10**	1.00

Note. Columns 1 and 2: Number and description of the functional categories according to the EcoCyc database. Categories of ontology level 1 are indicated in boldface, categories of level 2 in italics, and categories of level 3 in regular type. Column 3: *p*-values for the indirectly PhoPQ-regulated genes. Column 4: *p*-values for the directly PhoPQ-regulated genes. *p*-values describe the statistical significance of the enrichment of a functional class within the subset of PhoPQ-regulated genes. The most significant fields (*p*-value below 0.2) are boldfaced.

gene was shown to be PhoPQ dependent and important for the modification of lipid A (Guo et al. 1997), but in addition, our results point to its direct regulation. A third gene, *pqaB*, is shown to be involved in antimicrobial peptide resistance in *S. typhi* but is suggested to be indirectly PhoPQ regulated via the PmrAB system (Baker et al. 1999). This means that the detected PhoP box is either false positive or points toward a complex dual regulation of this gene.

Although not significantly enriched within the set of directly PhoPQ-regulated genes, the functional class related to pathogenicity and virulence also contained potential PhoPQ-dependent targets (*pagC, mgtC, virK*, and *STM0306*). Although the three former genes have previously been related to PhoPQ dependency, we find here, in addition, evidence for a potential direct dependency on PhoPQ. PagC is a membrane protein (Gunn et al. 1995) that was shown to be essential for survival within macrophages and for virulence in *S. typhimurium* (Pulkkinen and Miller 1991). MgtC was also shown to be required for intramacrophage survival and growth under low-Mg^{2+} conditions (Blanc-Potard and Groisman 1997). The exact function of MgtC, however, is still unknown. VirK contributes to the resistance of *S. typhimurium* against polymyxine B and is important for the systemic infection of the bacteria (Detweiler et al. 2003). Moreover, both *pagC* and *mgtC* are located on the SPI-3 pathogenicity island and were previously suggested to be acquired via horizontal gene transfer (Gunn et al. 1995; Blanc-Potard and Lafay 2003). This would imply that—after acquisition—both genes were integrated into the PhoPQ-dependent regulatory cascade of *S. typhimurium*. *STM0306*, a fourth PhoPQ-regulated gene that is involved in pathogenicity, is a paralog of the *S. typhimurium* SapA protein, which was shown to play a role in virulence (Parra-Lopez et al. 1993). An ortholog of SapA is present in *E. coli* and *Erwinia chrysanthemi*. In the latter, SapA was shown to play a role in the resistance of this organism to antimicrobial peptides (Lopez-Solanilla et al. 1998).

Discussion

Based on the combination of microarray and motif data, the PhoPQ-dependent regulon in both *E. coli* and *S. typhimurium* was reconstructed. This reconstruction is the best estimate of the true regulon composition that can be made at this stage, due to the presence of inherent variation commonly observed in microarray experiments, the uncertainty about the motif requirements of the PhoP motif model, and the restricted availability of experimental data. Note that, for instance, the *E. coli* and the *Salmonella* microarray experiments were not performed under

exactly the same conditions. However, we can expect that the conditions that were used trigger a large part of the PhoPQ regulon in both organisms. Indeed, the *E. coli* dataset tests the influence of Mg^{2+}, the most important PhoPQ signal. The *Salmonella* experiment, on the other hand, makes use of a constitutive mutant and therefore is relatively independent of the conditions applied. Conclusively, the size of the *E. coli* regulon might be underestimated compared to the size of the *S. typhimurium* regulon. This, however, does not prevent us from studying the overlap in regulon composition: a large overlap in regulon composition would imply that at least the targets induced by the major PhoPQ trigger, i.e., Mg^{2+} detected in *E. coli*, would be contained within the PhoPQ regulon in *Salmonella*. This not being the case, as we observed in this study, indicates that the true overlap indeed will be low.

Statistical analysis of the PhoPQ-related expression data of *S. typhimurium* clearly pointed out the pleiotropic nature of the PhoPQ regulatory system. We found evidence for at least 2855 genes being affected by the PhoP mutation. From these, only a limited subset of approximately 42 operons was directly regulated by PhoP, which is in accordance with previous predictions (Soncini et al. 1996; Soncini and Groisman 1996). The seemingly contradictory observations of the pleiotropic nature of the PhoPQ system, on the one hand, and the small size of the direct PhoPQ-dependent regulon, on the other hand, can be explained by the high number of regulatory proteins that are part of the direct regulon (e.g., *mig-14, slyA, pmrD, traM, STM0859*). Each of these regulators can activate other regulatory cascades (Lejona et al. 2003), allowing a combinatorial increase in PhoPQ-affected genes, e.g., *pmrD* encodes a protein that posttranscriptionally activates the two-component PmrAB system (Groisman 2001; Kato et al. 2003).

In an initial attempt to compare the overlap in the PhoPQ regulon composition between *E. coli* and *S. typhimurium*, we used a very strict definition of overlap: when two orthologs were differentially expressed in both organisms (data not shown). Although such a stringent analysis will result in a highly reliable set of potential PhoPQ targets, overlapping targets might escape detection because of potential type II error in the expression data (presence of false negatives in either one of the organisms). Therefore, we repeated the analysis using the low-stringent criteria described under Materials and Methods. Using this higher-sensitivity analysis, however, did not drastically increase the detected overlap in regulon composition, proving the biological relevance of the detected low overlap. Indeed, comparing the average regulon composition using nonstringent criteria, an overlap of only 13 PhoPQ regulated operons was observed (i.e., 26 genes).

For estimation of the direct regulon, we made approximations based on both stringent and nonstringent sequence requirements of the PhoP motif, assuming that the true sequence requirements must be somewhere in between. Besides the *mgtA* gene, which was missing in our analysis for reasons explained before, only 2 of these 13 overlapping operons (*phoPQ* and *slyB*) were directly regulated by the PhoPQ two-component system in both organisms.

In addition to these overlapping operons, both organisms had a considerable set of operons for which the PhoPQ dependency was confined to either one of the two organisms. These results point toward a high specialization of the PhoPQ regulon in either one of the species. This might not be so unexpected in view of the major role the PhoPQ system plays in determining the virulence phenotype of *S. typhimurium*, a phenotype that is absent in the related nonpathogenic strain *E. coli* K12 (Groisman 2001). This also explains why a large part of the direct regulon comprises genes that are unique for both species. These general conclusions seemed relatively independent of the stringency of the definition of "regulon overlap" and of the motif requirements used, indicating that the true overlap between the PhoPQ regulon compositions in both species is indeed low.

In *S. typhimurium*, significantly enriched functional classes within the set of directly PhoPQ-regulated genes are involved in "central intermediary metabolism," "synthesis and modification of cell envelope," "cation transport," and "drug sensitivity." The most significantly enriched classes (i.e., lowest *p*-values) in the subset of *E. coli* PhoPQ directly dependent genes are also related to "general metabolism" and "cell and membrane structure." Despite the limited overlap in regulon composition between the two species, the PhoPQ regulatory system seems to have conserved common functions in both species.

Conclusion

Our analysis shows how a regulatory system that is very well conserved (Kasahara et al. 1992) and that corresponds to the same extracellular signal can become integrated in a relatively short time period (120–160 million years (Cotter and DiRita 2000)) in seemingly completely different pathways. This acquisition of novel target genes might explain the high ability of prokaryotic organisms to evolve novel phenotypes and adapt to specific niches. The PhoPQ regulon might have recruited genes that contribute to a virulence phenotype in *S. typhimurium* but not in *E. coli* K12.

Acknowledgments. K. Marchal is a postdoctoral researcher of the Belgian Fund for Scientific Research (FWO-Vlaanderen). This work was partially supported by (1) IWT projects STWW-00162 and GBOU-SQUAD-20160; (2) Research Council KULeuven GOA Mefisto-666, GOA-Ambiorics, and IDO genetic networks; (3) FWO projects G.0115.01 and G.0413.03; and (4) IUAP V-22 (2002–2006). The authors like to thank Gert Thijs for the help with motif detection.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bader MW, Navarre WW, Shiau W, Nikaido H, Frye JG, Mc Clelland M, Fang FC, Miller SI (2003) Regulation of *Salmonella typhimurium* virulence gene expression by cationic antimicrobial peptides. Mol Microbiol 50:219–230

Baker SJ, Gunn JS, Morona R (1999) The *Salmonella typhi* melittin resistance gene; *pqaB* affects intracellular growth in PMA-differentiated U937 cells, polymyxin B resistance and lipopolysaccharide. Microbiology 145:367–378

Bearson BL, Wilson L, Foster JW (1998) A low pH-inducible, PhoPQ-dependent acid tolerance response protects *Salmonella typhimurium* against inorganic acid stress. J Bacteriol 180:2409–2417

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2003) GenBank. Nucleic Acids Res 31:23–27

Blanc-Potard AB, Groisman EA (1997) The *Salmonella selC* locus contains a pathogenicity island mediating intramacrophage survival. EMBO J 16:5376–5385

Blanc-Potard AB, Lafay B (2003) MgtC as a horizontally-acquired virulence factor of intracellular bacterial pathogens: evidence from molecular phylogeny and comparative genomics. J Mol Evol 57:479–486

Cotter PA, DiRita VJ (2000) Bacterial virulence gene regulation: an evolutionary perspective. Annu Rev Microbiol 54:519–565

Darwin KH, Miller VL (1999) Molecular basis of the interaction of *Salmonella* with the intestinal mucosa. Clin Microbiol Rev 12:405–428

De Smet F, Moreau Y, Engelen K, Timmerman D, Vergote I, De Moor B (2004) Balancing false positives and false negatives for the detection of differential expression in malignancies. Br J Cancer 91:1160–1165

Detweiler CS, Monack DM, Brodsky IE, Mathew H, Falkow S (2003) *virK*, *somA* and *rcsC* are important for systemic *Salmonella enterica* serovar *Typhimurium* infection and cationic peptide resistance. Mol Microbiol 48:385–400

Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30:1575–1584

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 USA 83:5189–5193

Fields PI, Groisman EA, Heffron F (1989) A *Salmonella* locus that controls resistance to microbicidal proteins from phagocytic cells. Science 243:1059–1062

Foster JW, Hall HK (1990) Adaptive acidification tolerance response of *Salmonella typhimurium*. J Bacteriol 172:771–778

Garcia VE, Soncini FC, Groisman EA (1996) Mg^{2+} as an extracellular signal: environmental regulation of *Salmonella* virulence. Cell 84:165–174

Gibbons HS, Lin S, Cotter RJ, Raetz CR (2000) Oxygen requirement for the biosynthesis of the S-2−hydroxymyristate moiety in *Salmonella typhimurium* lipid A. Function of LpxO, A new Fe^{2+}/alpha-ketoglutarate-dependent dioxygenase homologue. J Biol Chem 275:32940–32949

Groisman EA (2001) The pleiotropic two-component regulatory system PhoP–PhoQ. J Bacteriol 183:1835–1842

Groisman EA, Chiao E, Lipps CJ, Heffron F (1989) *Salmonella typhimurium phoP* virulence gene is a transcriptional regulator. Proc Natl Acad Sci USA 86:7077–7081

Groisman EA, Heffron F, Solomon F (1992a) Molecular genetic analysis of the *Escherichia coli phoP* locus. J Bacteriol 174:486–491

Groisman EA, Parra-Lopez C, Salcedo M, Lipps CJ, Heffron F (1992b) Resistance to host antimicrobial peptides is necessary for *Salmonella* virulence. Proc Natl Acad Sci USA 89:11939–11943

Guina T, Yi EC, Wang H, Hackett M, Miller SI (2000) A PhoP-regulated outer membrane protease of *Salmonella enterica* serovar *typhimurium* promotes resistance to alpha-helical antimicrobial peptides. J Bacteriol 182:4077–4086

Gunn JS, Alpuche-Aranda CM, Loomis WP, Belden WJ, Miller SI (1995) Characterization of *Salmonella typhimurium pagC/pagD* chromosomal region. J Bacteriol 177:5040–5047

Gunn JS, Lim KB, Krueger J, Kim K, Guo L, Hackett M, Miller SI (1998) PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. Mol Microbiol 27:1171–1182

Guo L, Lim KB, Gunn JS, Bainbridge B, Darveau RP, Hackett M, Miller SI (1997) Regulation of lipid A modifications by *Salmonella typhimurium* virulence genes *phoP–phoQ*. Science 276:250–253

Ideker T, Thorsson V, Siegel AF, Hood LE (2000) Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data. J Comput Biol 7:805–817

Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet 2:343–372

Karp PD, Arnaud M, Collado-Vides J, Ingraham J, Paulsen IT, Saier MH Jr (2004) The *E. coli* EcoCyc Database: No longer just a metabolic pathway database. ASM News 70:25–30

Kasahara M, Nakata A, Shinagawa H (1992) Molecular analysis of the *Escherichia coli phoP-phoQ* operon. J Bacteriol 174:492–498

Kato A, Tanabe H, Utsumi R (1999) Molecular characterization of the PhoP–PhoQ two-component system in *Escherichia coil* K-12: identification of extracellular Mg^{2+}-responsive promoters. J Bacteriol 181:5516–5520

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 USA 100:4706–4711

Lejona S, Aguirre A, Cabeza ML, Garcia VE, Soncini FC (2003) Molecular characterization of the Mg^{2+} responsive PhoP-PhoQ regulon in *Salmonella enterica*. J Bacteriol 185:6287–6294

Lopez-Solanilla E, Garcia-Olmedo F, Rodriguez-Palenzuela P (1998) Inactivation of the *sapA* to *sapF* locus of *Erwinia chrysanthemi* reveals common features in plant and animal bacterial pathogenesis. Plant Cell 10:917–924

Marchal K, Thijs G, De Keersmaecker SD, Monsieurs P, De Moor B, Vanderleyden J (2003) Genome-specific higher-order background models to improve motif detection. Trends Microbiol 11:61–66

Marchal K, De Keersmaecker S, Monsieurs P, van Boxel N, Lemmens K, Thijs G, Vanderleyden J, De Moor B (2004) In silico identification and experimental validation of PmrAB targets in *Salmonella typhimurium* by regulatory motif detection. Genome Biol 5:R9.1–R9.20

Miller SI, Mekalanos JJ (1990) Constitutive expression of the *phoP* regulon attenuates *Salmonella* virulence and survival within macrophages. J Bacteriol 172:2485–2490

Miller SI, Kukral AM, Mekalanos JJ (1989) A two-component regulatory system (*phoPphoQ*) controls *Salmonella typhimurium* virulence. Proc Natl Acad Sci USA 86:5054–5058

Minagawa S, Ogasawara H, Kato A, Yamamoto K, Eguchi Y, Oshima T, Mori H, Ishihama A, Utsumi R (2003) Identification and molecular characterization of the Mg^{2+} stimulon of *Escherichia coli*. J Bacteriol 185:3696–3702

Moreno-Hagelsieb G, Collado-Vides J (2002) A powerful non-homology method for the prediction of operons in prokaryotes. Bioinformatics 18:S329–S336

Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG (2001) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar *Typhi* CT18. Nature 413:848–852

Parra-Lopez C, Baer MT, Groisman EA (1993) Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in *Salmonella typhimurium*. EMBO J 12:4053–4062

Prouty AM, Gunn JS (2000) *Salmonella enterica* serovar *typhimurium* invasion is repressed in the presence of bile. Infect Immun 68:6763–6769

Pulkkinen WS, Miller SI (1991) A *Salmonella typhimurium* virulence protein is similar to a *Yersinia enterocolitica* invasion protein and a bacteriophage lambda outer membrane protein. J Bacteriol 173:86–93

Salgado H, Moreno-Hagelsieb G, Smith TF, Collado-Vides J (2000) Operons in *Escherichia coli*: genomic analyses and predictions. Proc Natl Acad Sci USA 97:6652–6657

Salgado H, Gama-Castro S, Martinez-Antonio A, Diaz-Peredo E, Sanchez-Solano F, Peralta-Gil M, Garcia-Alonso D, Jimenez-Jacinto V, Santos-Zavaleta A, Bonavides-Martinez C, Collado-Vides J (2004) RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. Nucleic Acids Res 32:D303–D306

Soncini FC, Groisman EA (1996) Two-component regulatory systems can interact to process multiple environmental signals. J Bacteriol 178:6796–6801

Soncini FC, Vescovi EG, Groisman EA (1995) Transcriptional autoregulation of the *Salmonella typhimurium phoPQ* operon. J Bacteriol 177:4364–4371

Soncini FC, Garcia VE, Solomon F, Groisman EA (1996) Molecular basis of the magnesium deprivation response in *Salmonella typhimurium*: identification of PhoP-regulated genes. J Bacteriol 178:5092–5099

Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. Proc Natl Acad Sci USA 100:9440–9445

Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999) Systematic determination of genetic network architecture. Nat Genet 22:281–285

Thijs G (2003) Probabilistic methods to search for regulatory elements in sets of coregulated genes. PhD thesis. Faculty of Applied Sciences, Katholieke Universiteit Leuven, Leuven

Thijs G, Lescot M, Marchal K, Rombauts S, De Moor B , Rouze P, Moreau Y (2001) A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. Bioinformatics 17:1113–1122

Thijs G, Marchal K, Lescot M, Rombauts S, De Moor B , Rouze P, Moreau Y (2002) A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes. J Comput Biol 9:447–464

van Velkinburgh JC, Gunn JS (1999) PhoP–PhoQ-regulated loci are required for enhanced bile resistance in *Salmonella* spp. Infect Immun 67:1614–1622

Vescovi EG, Ayala YM, Di Cera E, Groisman EA (1997) Characterization of the bacterial sensor protein PhoQ. Evidence for distinct binding sites for Mg^{2+} and Ca^{2+}. J Biol Chem 272:1440–1443

Yamamoto K, Ogasawara H, Fujita N, Utsumi R, Ishihama A (2002) Novel mode of transcription regulation of divergently overlapping promoters by PhoP, the regulator of two-component system sensing external magnesium availability. Mol Microbiol 45:423–438