Genome-wide mapping of the *Escherichia coli* PhoB regulon reveals many transcriptionally inert, intragenic binding sites

Devon Fitzgerald[1,2]*, Anne Stringer[1], Carol Smith[1], Pascal Lapierre[1], and Joseph T. Wade[1,2,3]

[1]Wadsworth Center, New York State Department of Health, Albany, New York, USA.

[2]Department of Biomedical Sciences, School of Public Health, University at Albany, Albany, New York, USA.

[3]Corresponding author: joseph.wade@health.ny.gov

*Current address: TwinStrand Biosciences, Seattle, Washington, USA.

ABSTRACT

Genome-scale analyses have revealed many transcription factor binding sites within, rather than upstream of genes, raising questions as to the function of these binding sites. Here, we use complementary approaches to map the regulon of the *Escherichia coli* transcription factor PhoB, a response regulator that controls transcription of genes involved in phosphate homeostasis. Strikingly, the majority of PhoB binding sites are located within genes, but these intragenic sites are not associated with detectable transcription regulation and are not evolutionarily conserved. Many intragenic PhoB sites are located in regions bound by H-NS, likely due to shared sequence preferences of PhoB and H-NS. However, these PhoB binding sites are not associated with transcription regulation even in the absence of H-NS. We propose that for many transcription factors, including PhoB, binding sites not associated with promoter sequences are transcriptionally inert, and hence are tolerated as genomic “noise”.

IMPORTANCE

Recent studies have revealed large numbers of transcription factor binding sites within the genes of bacteria. The function, if any, of the vast majority of these binding sites has not been investigated. Here, we map the binding of the transcription factor PhoB across the *Escherichia coli* genome, revealing that the majority of PhoB binding sites are within genes. We show that PhoB binding sites within genes are not associated with regulation of the overlapping genes. Indeed, our data suggest that bacteria tolerate the presence of large numbers of non-regulatory, intragenic binding sites for transcription factors, and that these binding sites are not under selective pressure.

INTRODUCTION

Bacterial transcription factors often bind sites within genes

Bacteria encode numerous transcription factors (TFs) that regulate transcription initiation by binding DNA near promoters and modulating the ability of RNAP holoenzyme to bind promoter DNA or to isomerize to an actively transcribing conformation (1). TF function has been studied almost exclusively in the context of TF binding sites in intergenic regions, upstream of the regulated genes. However, genome-scale analyses of TF binding have identified large numbers of intragenic binding sites, far from gene starts. The proportion of binding sites for a TF that are intragenic varies extensively between different TFs (2, 3), with some TFs having the majority of their binding sites inside genes (3–6). Despite the large number of intragenic TF binding sites, relatively little is known about their function.

Regulatory activity has been described for few intragenic TF binding sites, and can be classified into distinct classes based on the regulatory target and mechanism of action: (i) canonical regulation of transcription initiation of the downstream gene, generating an RNA with an extended 5' UTR that overlaps a gene (5, 7–10); (ii) canonical regulation of transcription initiation of a stable non-coding RNA that initiates inside a gene or 3' UTR (11, 12); (iii) regulation of transcription initiation of the gene that contains the TF binding site; mechanisms of regulation in almost all such cases are unknown (3), although transcription repression can occur from a site close to the promoter due to a physical interaction with a more upstream site, resulting in formation of a DNA loop (13, 14); (iv) regulation of transcription elongation due to the TF acting as a road-block for RNAP (15–18). Another possible regulatory function for intragenic TF binding sites is regulation of pervasive transcription – transcription of large numbers of short, unstable RNAs from inside genes that is ubiquitous in bacteria (19, 20). Although there are no described examples of TFs that regulate unstable, intragenic transcripts, many of these RNAs are differentially expressed between growth conditions (21), consistent with regulation by TFs. Intragenic TF binding sites might also have functions that are not directly connected to gene regulation,

such as facilitating short- or long-range chromosome contacts (22–25), or serving as TF-titrating decoy sites (26, 27). Lastly, it is possible that intragenic TF binding sites serve no biological function, and arise as a consequence of genetic drift (28), or genome evolution that is constrained by selection for particular codons.

PhoB is a conserved transcription factor that regulates phosphate homeostasis

PhoB is a member of the PhoB/OmpR family of response regulator TFs, and is a key regulator of phosphate homeostasis in many Gram-negative bacteria (29, 30). PhoB forms a two-component system with the sensor kinase PhoR (31). When inorganic phosphate (P_i) levels are low, PhoR autophosphorylates and then phosphorylates PhoB (30, 31), triggering PhoB dimerization and DNA binding activity (30). Phosphorylated PhoB binds direct repeat sequences called *pho* boxes (32), and is a dual regulator, capable of both activating and repressing transcription depending on the position of the binding site.

In *Escherichia coli* and related species, PhoB regulates expression of genes encoding the high affinity phosphate transport system (*pst*), a phosphonate transport complex (*phn*), the glycerol-3-phosphate transporter (*ugp*), and other genes related to phosphate homeostasis (30, 33). These genes are collectively referred to as the *pho* regulon. PhoB has been implicated in regulation of a number of other cellular processes and stress responses, including motility, biofilm formation, quorum sensing, cell surface remodeling, stringent response, and the general stress response (34, 35). Indeed, transcriptomic and proteomic studies of phosphate-depleted *E. coli* have suggested that the *pho* regulon has many additional members (36, 37). However, most of these putative regulon members have limited experimental support (30, 33).

Here, we describe a high-resolution, genome-wide mapping of the *pho* regulon using ChIP-seq and RNA-seq. We refine and expand the set of known *pho* regulon genes and identify many intragenic PhoB binding sites. We show that the large majority of intragenic PhoB binding sites are not conserved, and are not associated with

detectable regulatory function. Thus, our data suggest that individual intragenic PhoB sites are non-functional, and that TFs can bind many intragenic sites with little or no impact on local transcription.

RESULTS

Genome-wide Binding of PhoB under phosphate-limiting conditions

ChIP-seq is used to map the genome-wide binding of TFs. To facilitate ChIP-seq of *E. coli* PhoB, we introduced C-terminal FLAG tags at the native *phoB* locus. We used quantitative reverse-transcriptase PCR (qRT-PCR) to measure expression of *pstS*, a PhoB-activated gene, in wild-type cells, Δ*phoB* cells, and cells expressing *phoB*-FLAG$_3$. Cells were grown in minimal medium with low phosphate levels, to induce the kinase activity of PhoR. As expected, we observed a large decrease (~900-fold) in *pstS* levels in Δ*phoB* cells relative to wild-type cells (Figure 1). In cells expressing PhoB-FLAG$_3$, we observed a much smaller decrease (~8-fold) in *pstS* levels relative to wild-type cells (Figure 1), indicating that the tagged PhoB derivative retains partial function.

We used ChIP-seq to map the genome-wide binding of PhoB-FLAG$_3$ during growth under low phosphate conditions. Thus, we identified 65 enriched regions (Figure 2A; Table 1). As a control, we performed ChIP-seq with an untagged strain grown under the same conditions; none of the regions enriched in the PhoB-FLAG$_3$ ChIP-seq dataset were enriched in the control dataset (Figure 2A). We conclude that the 65 enriched regions in the PhoB-FLAG$_3$ ChIP-seq dataset are likely to represent genuine PhoB-bound regions. Note that a single PhoB-bound region could include more than one PhoB site, as is the case for the region upstream of *phoB* itself, which has been reported to include two PhoB sites (38). We identified a highly enriched sequence motif, with instances of the motif found in 59 of the 65 putative PhoB-bound regions (MEME *E*-value = 3.0e^{-72}; Figure 2B). This motif contains a clearly distinguishable direct repeat and is similar to the previously reported *pho* box consensus sequence (39). Furthermore, the identified motif is centrally enriched relative to the calculated ChIP-seq peak centers (Figure 2C; Centrimo *E*-value = 1.2e^{-12}). The presence and central enrichment of this motif at ChIP-seq peaks further supports the veracity of PhoB-bound regions and confirms the high spatial resolution of the ChIP-seq data.

The 65 PhoB-bound regions identified by ChIP-seq include most well-established PhoB sites, as well as many novel targets (Table 1). We identified PhoB-bound regions upstream of 7 of the 10 genes/operons described previously as being in the *pho* regulon (Table 1, underlined gene names) (33), with no ChIP-seq signal upstream of *waaH*, *ytfK*, or *psiE*. We also identified PhoB-bound regions upstream of the predicted *pho* regulon gene *amn* (Table 1) (40), and upstream of *yoaI*, which was described as a direct PhoB target in *E. coli* O157:H7 (41, 42). We identified 21 PhoB-bound regions upstream of genes/operons not previously described as part of the *pho* regulon, and lacking a clear connection to phosphate homeostasis. The remaining 36 PhoB-bound regions, over half of the total sites identified by ChIP-seq, are located inside genes (Figure 2D; Table 1). Strikingly, all but 3 intragenic PhoB binding sites are far from neighboring gene starts (>200 bp), thus are unlikely to participate in promoter-proximal regulation of these genes (Figure 2D; Table 1).

Our PhoB ChIP-seq data show only modest agreement with an earlier study that identified many putative PhoB binding sites using ChIP-chip (43), although both studies are consistent in the lack of signal upstream of *waaH*, *ytfK*, or *psiE*. Of the 43 ChIP-chip peaks identified by Yang *et al.*, 24 are <400 bp from a ChIP-seq peak in our data (Table 1, coordinates in red), while the remaining 19 ChIP-chip peaks are >2,800 bp from the closest ChIP-seq peak. Even for the 24 ChIP-chip peaks close to ChIP-seq peaks identified in the current study, the peak centers calculated from the two datasets are up to 383 bp apart, and only 13 regions share a motif call between studies (Table 1, motifs in red). These discrepancies between datasets are probably due, at least in part, to the low resolution of the ChIP-chip data, and differences in peak-calling and motif-calling algorithms. It is challenging to determine whether peak calls from each of the two studies represent the same biological binding events. We performed *de novo* motif identification for three sets of peaks: (i) shared, (ii) unique to the current study, and (iii) unique to Yang *et al*. For (i) shared peaks and (ii) peaks unique to the current study, 100 bp of sequence surrounding each ChIP-seq peak center was extracted and analyzed by MEME. In both cases, highly enriched sequence motifs were found that are close matches to the expected PhoB motif (Figure 3A + B). For (iii) sites unique to the Yang *et al.* dataset, the same analysis was performed using both 100 bp and 500 bp

windows surrounding the published peak center locations. The resulting motifs were poorly enriched (MEME *E*-values >1) and bear no similarity to the expected PhoB motif. We conclude that most or all of the 40 regions unique to the current study represent genuine PhoB binding sites, while those unique to the Yang *et al.* study largely do not.

Genome-wide Binding of PhoB under high phosphate conditions

To determine whether PhoB binds any target DNA sites when PhoR is inactive, we repeated the ChIP-seq experiment, but grew cells under conditions with high phosphate. We detected only a single PhoB-bound region: the intergenic region upstream of *pstS* (Figure 2A). This is a well-established site of PhoB binding, and was the most enriched PhoB-bound region in the low phosphate ChIP-seq experiment. As expected, PhoB binding upstream of *pstS* was substantially lower under conditions of high phosphate than under conditions of low phosphate (Figure 2A). Thus, our data suggest that under conditions of high phosphate, PhoB weakly regulates *pstS*, but does not regulate any of its other target genes.

Reassessing the pho *regulon*

To address whether the detected PhoB sites contribute to transcription regulation, RNA-seq was performed using wild-type and Δ*phoB* strains grown in low-phosphate medium. In total, 181 genes were differentially expressed between the wild-type and the Δ*phoB* strains (p-value ≤ 0.01, >2-fold difference in RNA levels; Figure 4; Table S1). We observed significant positive regulation of all 7 reported *pho* regulon operons for which we observed upstream PhoB binding by ChIP-seq, i.e., *phnCDEFGHIJKLMNOP*, *phoH*, *ugpBAECQ*, *pstSCAB-phoU*, *phoA-psiF*, *phoE*, and *phoBR* (Table 1 + S1; positive regulation was observed for all genes in all operons, except for *phoB* which could not be assessed in the Δ*phoB* strain). We also observed significant positive regulation of *amn* and *yoaI*; ChIP-seq identified PhoB binding upstream of these genes, and although they have not generally been considered as part of the *pho* regulon, they have been previously reported as being direct PhoB targets (41, 42).

We observed significant positive regulation of *ytfK* and *waaH*, reported *pho* regulon genes that lack associated PhoB binding. We conclude that *ytfK* and *waaH* are regulated indirectly by PhoB. By contrast, we did not observe significant regulation of known and predicted *pho* regulon genes *psiE*, *asr*, *eda*, *argP*, and *pitB*; none of these genes had detectable upstream PhoB binding by ChIP-seq. We conclude that *psiE*, *asr*, *eda*, *argP*, and *pitB* are unlikely to be regulatory targets of PhoB.

To identify novel *pho* regulon genes, we determined whether any additional genes with associated PhoB binding were significantly differentially expressed between wild-type and Δ*phoB* cells. We observed significant, >2-fold positive regulation of *cusC* and *yibT*, and significant, >2-fold negative regulation of *feaR* and *agp*, genes with upstream PhoB binding as determined by ChIP-seq. We also observed significant regulation of 6 genes with internal PhoB sites: *yahA*, *gloC*, *pnp*, *evgS*, *eptB* and *malF*. Although only two of these genes (*yahA* and *evgS*) were differentially expressed >2-fold between wild-type and Δ*phoB* cells, all but *yahA* were more highly expressed in Δ*phoB* cells than wild-type cells. We hypothesized that PhoB represses transcription of these genes by acting as a roadblock for RNAP. To test this hypothesis, we grew wild-type and Δ*phoB* cells under phosphate-limiting conditions and measured RNAP (β subunit) occupancy upstream and downstream of the PhoB sites within the *gloC*, *pnp*, and *evgS* genes using ChIP-qPCR. As controls, we measured RNAP occupancy within the *pstS*, and *ugpB* genes, confirmed members of the *pho* regulon. We also measured RNAP occupancy within the *yoaI* and *amn* genes that the combined RNA-seq and ChIP-seq data suggested are members of the *pho* regulon. As expected, we observed substantially higher RNAP occupancy within *pstS* and *ugpB* in wild-type cells than in Δ*phoB* cells (Figure 5). Moreover, we observed substantially higher RNAP occupancy within *yoaI* and *amn* in wild-type cells than in Δ*phoB* cells (Figure 5), supporting the idea that these genes are part of the *pho* regulon. For the three genes with intragenic PhoB sites, we reasoned that if PhoB acts as a roadblock to elongating RNAP, RNAP occupancy downstream of PhoB sites would increase in Δ*phoB* cells relative to that in wild-type cells and relative to any change in RNAP occupancy

upstream of the PhoB site. However, we did not observe significant increases in relative RNAP occupancy downstream of PhoB sites for any of the three genes (Figure 5; $p > 0.2$; see Methods for details of the statistical analysis), suggesting that PhoB regulation of these genes is indirect.

PhoB-dependent recruitment of initiating RNAP

The majority of the PhoB binding sites identified by ChIP-seq were not associated with regulation detectable by RNA-seq. We hypothesized that this could be due to three reasons: (i) the binding sites are non-regulatory, (ii) regulation is condition-specific and/or requires additional factors, or (iii) PhoB regulates transcription of short, unstable, non-coding RNAs that are not detectable by conventional RNA-seq. To test the latter possibility, we used ChIP-seq to measure the association of σ^{70} in regions close to PhoB binding sites. σ^{70} is rapidly released from RNAP upon the transition from transcription initiation to elongation (44); thus, σ^{70} occupancy on DNA, as measured by ChIP-seq, is an indication of the level of association of initiating RNAP with DNA. Since transcription initiation occurs prior to RNA processing, σ^{70} occupancy can be observed even at promoters of highly unstable RNAs (45).

To measure the effects of PhoB on RNAP holoenzyme recruitment, we performed ChIP-seq of σ^{70} in wild-type and Δ*phoB* strains grown in low-phosphate medium. Normalized σ^{70} occupancy was calculated for 400 bp windows surrounding each PhoB binding site to systematically assess σ^{70} binding at these sites (Figure 6). Three PhoB binding sites showed large reductions (>19-fold) in σ^{70} occupancy in the Δ*phoB* strain relative to wild-type. Two of these sites are associated with the *phoB* gene itself; σ^{70} occupancy measurements at these sites are impacted by the loss of associated DNA sequence resulting from deletion of *phoB*. The third PhoB site is the regulatory site upstream of *pstS*. We conclude that PhoB activates *pstS* transcription at the level of RNAP recruitment, as suggested by structural models of the DNA:PhoB:RNAP complex (46–48). PhoB binding sites upstream of *yoaI*, *phoA*, *mglB*, *phnC*, and *phoH*, showed >2-fold lower σ^{70} occupancy in the Δ*phoB* strain relative to wild-type, suggesting that PhoB recruits initiating RNAP to these promoters. These data are largely

consistent with the RNA-seq data showing >2-fold differential expression of *yoaI*, *phoA*, *phnC* and *yoaI* between wild-type and Δ*phoB* cells (Table 1). For most other PhoB sites, including almost all intragenic sites, σ^{70} occupancy was low in both wild-type and Δ*phoB* strains (Figure 6), strongly suggesting that these sites are not associated with active promoters under the growth conditions used. The remaining sites were associated with substantial σ^{70} occupancy that was similar in both wild-type and Δ*phoB* strains, suggesting that they are close to active promoters whose activity is independent of PhoB under the conditions tested.

H-NS co-associates with many intragenic PhoB sites, but does not block RNAP recruitment

We noted that 18 PhoB sites (12 intragenic and 6 intergenic), representing 28% of all sites identified by ChIP-seq, are in regions bound by the nucleoid-associated protein H-NS (49). Thus, PhoB sites are significantly enriched in H-NS-bound regions, which only represent 17% of the genome (Binomial test $p = 0.02$). Since H-NS is known to silence transcription (50), we hypothesized that the lack of detectable PhoB-dependent regulation at some sites may be due to the silencing effects of H-NS. To test this hypothesis, we repeated the σ^{70} ChIP-seq experiment in Δ*hns* and Δ*hns* Δ*phoB* strains. Comparison of σ^{70} occupancy between wild-type and Δ*hns* strains revealed substantially increased occupancy around some PhoB binding sites in the Δ*hns* strain, with most of these sites being intragenic (Figure 7A). Indeed, we observed widespread increases in σ^{70} occupancy at promoters genome-wide in the Δ*hns* strain relative to wild-type; most of the promoters showing increased σ^{70} association are located in regions of high H-NS occupancy (Figure 7B) (49). These data are consistent with our earlier study showing widespread transcriptional silencing by H-NS, particularly within genes (45). We next compared σ^{70} occupancy around PhoB binding sites between Δ*hns* and Δ*hns*Δ*phoB* strains (Figure 8). As for *hns*$^+$ cells, the only large differences (>5-fold) in σ^{70} occupancy were associated with the PhoB sites at *phoB* and *pstS*. Interestingly, we did not observe differences >1.5-fold in σ^{70} occupancy at any other sites, including the sites upstream of *yoaI*, *phoA*, *mglB*, *phnC*, and *phoH*. However, the differences in σ^{70} occupancy we observed for *phoB* and *pstS* sites were between 2- and 4-fold lower than differences observed in the *hns*$^+$ strains at the same sites. Hence, the more subtle differences in σ^{70} occupancy observed at the sites

upstream of *yoaI*, *phoA*, *phnC*, and *phoH* in the *hns*$^+$ strains may have escaped detection in the Δ*hns* strains. The lower effect of PhoB on σ^{70} occupancy in the Δ*hns* strain background may be due to the large-scale redeployment of RNAP that occurs in the absence of H-NS (51).

While PhoB sites are enriched within H-NS-bound regions, H-NS does not appear to modulate PhoB activity at any site. We hypothesized that the enrichment of PhoB binding within H-NS-bound regions is due simply to the nucleotide content of the PhoB binding site; like H-NS-bound regions, the PhoB binding site has a higher A/T-content than the genome as a whole. To test this hypothesis, we scrambled the sequence of every PhoB binding site identified by ChIP-seq. We then derived a position weight matrix (PWM) from these scrambled sites and scored every genomic sequence for a match to this PWM. Strikingly, 36% of the top 1,000 scoring positions are within regions bound by H-NS (49). We conclude that the enrichment of PhoB binding sites within H-NS-bound regions is likely due to the A/T-rich nature of the binding motif.

Sequence conservation of PhoB binding sites

Sequence conservation of a DNA binding site is often an indication that the site is functional (52). We determined the sequence conservation of the 59 *E. coli* PhoB sites identified by ChIP-seq for which we could identify an instance of the PhoB binding motif. Specifically, we scored homologous regions from 29 diverse γ-proteobacterial species for matches to the PhoB binding site motif (Figure 2C). The DNA-binding domain of PhoB is highly conserved across these species (Figure S1). As shown in Figure 9, the PhoB binding sites upstream of *pstS*, *phoB*, *phoA*, and *ugpB*, are broadly conserved. The PhoB binding sites upstream of *phoE* and *phoH* are conserved, albeit to a lesser degree. The PhoB binding site upstream of *phnC* is conserved in only a few species, suggesting that *phnC* is not a core member of the *pho* regulon. Among the novel PhoB binding sites, the best conserved is the site upstream of *rmf*, with strong matches to the PhoB DNA binding motif found upstream of *rmf* in most species analyzed. We detected PhoB upstream of *rmf* by ChIP-seq (Table 1) but did not detect significant PhoB-dependent regulation at the level of RNA abundance or RNAP recruitment. PhoB

binding sites upstream of *agp*, *rpoH*, *cusR*/*cusC*, and *yoaI* were also conserved, albeit to a less degree, similar to sites upstream of *phoE* and *phoH*. The remaining intergenic PhoB sites are not well conserved, with most having few or no strong matches to the PhoB DNA binding motif outside of *E. coli*. Lastly, we examined conservation of intragenic PhoB sites. In most cases, these sites have little or no conservation; however, PhoB sites within *flhD*, *phoB*, and *pnp* are conserved among roughly half the species examined. This conservation may reflect a conserved function for the PhoB binding site, or could be due to sequence constraints on the codons.

DISCUSSION

Comprehensive reassessment of the pho *regulon in* E. coli *and beyond*

By combining ChIP-seq and RNA-seq, we are able to reassess the *pho* regulon, with high resolution assignment of PhoB binding sites. As described above, the sensitivity and resolution of an earlier ChIP-chip study were substantially lower, precluding a comprehensive reassessment of the *pho* regulon (43). Previous studies disagree on which genes comprise the *pho* regulon in *E. coli*; however, most studies agree that the *pho* regulon includes the following operons: *pstSCAB-phoU*, *phoA*, *phoH*, *phnCDEFGHIJKLMNOP*, *phoBR*, *phoE*, *ytfK*, *ugpBAECQ*, and *psiE* (30, 33); *waaH* is considered in some studies to be in the *pho* regulon (33). Our data are largely consistent with these assignments, but provide strong evidence against *ytfK*, *psiE*, and *waaH* being members of the *pho* regulon. Specifically, we did not detect PhoB binding near any of these genes (Table 1), and we did not detect PhoB-dependent regulation of *psiE* (Table S1). It is formally possible that the C-terminal FLAG tags on PhoB altered its binding specificity or reduced its affinity for DNA such that we failed to detect binding to sites upstream of *ytfK*, *psiE*, and *waaH*. Nonetheless, in such a scenario, these binding sites would presumably have relatively low affinity for PhoB.

Our data also rule out several other putative *pho* regulon genes: *asr*, *eda*, *argP*, and *pitB* that were not associated with detectable PhoB binding or PhoB-dependent regulation (40, 53–57). Similarly, we did not detect binding of PhoB upstream of the sRNA-encoding gene *esrL*, despite a recent report of PhoB binding to this region in enteropathogenic *E. coli*, with the sequence of the reported PhoB site being identical in *E. coli* K-12 (58). By contrast, our ChIP-seq data support the assignment of *amn* and *yoaI* as *pho* regulon members, as has previously been suggested based on limited experimental evidence (40–42). Our ChIP-seq and RNA-seq data identify novel *pho* regulon members with confidence: *cusC*, *feaR*, *yibT*, and *agp*. These genes all have PhoB binding sites upstream, and were significantly, differentially expressed >2-fold between wild-type and Δ*phoB* strains. We note that direct positive regulation of *cusC* and direct negative regulation of *feaR* by PhoB have

been suggested previously (43). Our data provide no evidence to suggest that there are unannotated transcripts regulated by PhoB, or transcripts whose regulation by PhoB is masked by H-NS (Figures 6 + 8). Lastly, our data do not support direct regulation by PhoB sites located within genes (Figure 5).

Phylogenetic analysis of PhoB binding sites highlights a highly conserved set of *pho* regulon genes within the γ-proteobacteria: *pstS*, *phoB*, *phoA*, *ugpB*, and associated operonic genes (Figure 9). Consistent with this, direct PhoB regulation of the *pstS* and *phoB* transcripts has been described for the more distantly related α-proteobacterium *Caulobacter crescentus* (59). *phoE*, *phoH*, and *yoaI* represent a second set of conserved *pho* regulon genes, although their conservation is more phylogenetically restricted. Interestingly, while PhoB regulation of *phnC* and associated operonic genes does not appear to be widely conserved, we did observe evidence for strong PhoB sites upstream of *phnC* in a small set of species, and *phnC* is known to be a direct regulatory target of PhoB in *C. crescentus* (59), suggesting that *phnC* may have a niche-specific function in phosphate homeostasis.

The phylogenetic pattern of PhoB binding site conservation for sites upstream of *rmf*, *agp*, *rpoH*, and *cusR*/*cusC* suggests these genes may be part of the conserved *pho* regulon. We observed significant differential expression of *cusC* and *agp* between wild-type and Δ*phoB* cells. By contrast, we did not observe significant differential expression of *rmf* or *rpoH*. We speculate that regulation of *rmf* and *rpoH* by PhoB is integrated with regulation by other transcription factors, such that PhoB-dependent changes in expression are only detectable under specific growth conditions. Consistent with this idea, transcription of *rmf* has been shown to be regulated by ppGpp (60, 61) and CRP (62), and possibly by additional transcription factors (63) and diverse stress conditions (64). PhoB binding sites upstream of *agp*, *rpoH*, and *cusR*/*cusC* are conserved in largely the same set of species as binding sites upstream of *phoE*, *phoH*, and *yoaI*, suggesting that these species share a common set of *pho* regulon genes.

Most intragenic PhoB sites appear to be non-functional, and are not under selective pressure

Our data argue against intragenic PhoB sites having regulatory activities of the types that have been described previously for intragenic TF sites; specifically, regulation of transcription from an intragenic promoter (5, 7–12), or regulation of the overlapping gene either by roadblock repression or a novel mechanism (3, 15–18). Indeed, most intragenic PhoB sites are associated with little or no local σ^{70} binding (Figure 6), indicating that PhoB binding alone is insufficient to recruit RNAP. Thus, it is likely that RNAP:σ^{70}-interacting promoter elements are also necessary for PhoB-dependent recruitment of RNAP. Moreover, the spacing between PhoB sites and core promoter elements is likely to be important in determining whether PhoB recruits RNAP, since even intragenic PhoB sites that are close to intragenic promoters (i.e., those associated with high ChIP-seq signal for σ^{70}) show no difference in σ^{70} occupancy upon deletion of *phoB*. Consistent with this idea, structural models of the PhoB:RNAP:DNA complex formed at PhoB-activated promoters support strict spacing requirements between the *pho* box and core promoter elements (46–48, 65).

Widespread intragenic TF binding is emerging as a common phenomenon as more TFs are mapped using ChIP-seq (2–6, 66). Similarly, many σ factors have been shown to bind and initiate transcription from large numbers of intragenic promoters (8, 19, 20, 67–70). In the majority of cases tested, these intragenic binding sites are poorly conserved (69, 71), as is the case for intragenic PhoB sites (Figure 9). Based on the lack of detectable transcriptional activity, and the limited conservation of intragenic PhoB sites, we speculate that intragenic TF sites often arise due to genetic drift or selective pressures on overlapping sequences such as codons. Consistent with this idea, intragenic PhoB sites tend to be weaker (lower ChIP-seq enrichment) than intergenic sites (Mann-Whitney U test, $p = 0.005$). A previous study showed that the predicted number of intragenic binding sites for many bacterial TFs is the same in actual genome sequences as it is in randomized genome sequences, suggesting that intragenic TF binding sites are common and arise largely due to genetic drift (28). We further speculate that the fitness cost of intragenic PhoB sites is low. Intragenic TF binding sites can therefore be considered genomic "noise". Interestingly, the vast majority of PhoB binding events detected in *C. crescentus* are intergenic (59), suggesting that intragenic PhoB binding in *C. crescentus* may be associated with a fitness

cost. Finally, we cannot rule out the possibility that intragenic PhoB sites in *E. coli* are functional. For example, they could contribute, *en masse*, to titration of PhoB, they could facilitate DNA looping that impacts chromosome structure, as has been suggested for some TFs (22–27), or they could regulate transcription by unknown mechanisms.

We have comprehensively mapped the PhoB regulon by assessing PhoB binding, PhoB-dependent transcriptome changes, and PhoB-dependent RNAP recruitment. We identified novel *pho* regulon members, some of which are modestly conserved across other genera, and identified many seemingly non-functional PhoB binding sites inside genes. We conclude that a combination of binding site information (e.g., ChIP-seq) and regulatory information (e.g., RNA-seq) is required to accurately define the regulons of most TFs.

MATERIALS AND METHODS

Strains and plasmid

E. coli MG1655 and its derivatives were used for this study. The strains and plasmid used are listed in Table 2. All oligonucleotides used are listed in Table S2. For ChIP-seq, PhoB was C-terminally epitope tagged with a 3x-FLAG tag. The tag was inserted at the native *phoB* locus using FRUIT recombineering (72) using oligonucleotides JW2973 and JW2974. MG1655 Δ*phoB* (DMF84) and Δ*hns* (AMD565a) strains were constructed by P1 transduction from Keio collection strains (73) into MG1655. The kan^R genes were removed by FLP-recombinase expressed from pCP20, as previously described (74). The MG1655 Δ*hns* Δ*phoB* strain was made in the same manner, with deletions introduced sequentially. MG1655 Δ*phoB* (CDS091) was constructed using FRUIT (72) using oligonucleotides JW6280, JW6281, JW6294 and JW6295. Note that there are two MG1655 Δ*phoB* strains used in this study. DMF34 was used for σ^{70} ChIP-seq experiments, and CDS091 was used for qRT-PCR, RNA-seq, and RNAP ChIP-qPCR experiments.

Quantitative reverse transcription PCR

Quantitative RT-PCR was performed based on a previous study (69). MG1655 + pBAD24, CDS091 (MG1655 Δ*phoB*) + pBAD24, or DMF34 + pBAD24 cells were grown at 37° C with aeration in MOPS minimal medium with 0.2 mM K_2PO_4, 0.4% glucose and 100 µg/ml ampicillin to an OD_{600} of 0.5-0.6. Arabinose was added to a final concentration of 0.2% for 7 minutes or 20 minutes, with one or two of the three replicate samples for each strain receiving arabinose for 7 minutes. Thus, the replicate samples were not always consistent with respect to the extent of growth after arabinose addition. However, since arabinose is not expected to affect *pstS* expression, all samples for a single strain were treated as replicates regardless of whether cells were grown for 7 minutes or 20 minutes after arabinose addition. RNA was prepared as described for RNA-seq. RNA was reverse-transcribed using SuperScript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions. A control reaction omitting reverse transcriptase was performed. 1% of the cDNA (or negative

control) was used as a template in a quantitative real time PCR using an Applied Biosystems 7500 Fast real time PCR machine, with primers JW156 + JW157 for amplifying the *minD* control gene, and JW7802 + JW7803 for amplifying *pstS*. Relative expression of *pstS* was determined by the ΔC_T method, normalizing to *minD* expression.

ChIP-seq

For low phosphate growth experiments, cells were grown at 37° C with aeration in MOPS minimal medium with 0.2 mM K_2HPO_4 and 0.4% glucose, as previously described (37, 75). For high phosphate growth experiments, cells were grown in MOPS minimal medium with 1.32 mM K_2HPO_4 and 0.4% glucose. Subcultures were inoculated 1:100, and grown at 37° C with aeration to an OD_{600} of 0.5-0.7. ChIP-seq libraries were prepared as previously described (76). Libraries were prepared from two biological replicate cultures for each experimental group. For DMF34 (MG1655 phoB-$FLAG_3$) or MG1655, PhoB-$FLAG_3$ (or the untagged control) was immunoprecipitated with 2 μL of α-FLAG M2 monoclonal antibody (Sigma). For MG1655, DMF84 (Δ*phoB*), AMD565a (MG1655 Δ*hns*), and DMF85 (MG1655 Δ*phoB* Δ*hns*), σ^{70} was immunoprecipitated with 1 μL of α-σ^{70} antibody (Neoclone). Libraries were sequenced on a HiSeq 2000 (Illumina) by the University at Buffalo Next-Generation Sequencing Core Facility or a NextSeq (Illumina) by the Wadsworth Center Applied Genomic Technologies Core Facility.

Analysis of PhoB-$FLAG_3$ ChIP-seq Data

Duplicate ChIP-seq data for (i) MG1655 PhoB-$FLAG_3$ (DMF34) grown in low phosphate conditions, (ii) untagged MG1655 grown in low phosphate conditions, and (iii) MG1655 PhoB-$FLAG_3$ (DMF34) grown in high phosphate conditions, were aligned to the *E. coli* MG1655 genome (NC_000913.3) using CLC Genomics Workbench (version 8). ChIP-seq peaks were called using a previously described analysis pipeline (8).

Analysis of RNAP occupancy around PhoB binding sites

Wild-type MG1655 and MG1655 Δ*phoB* (CDS091) cells were grown at 37° C with aeration to an OD_{600} of 0.6-0.7 in MOPS minimal medium with 0.2 mM K_2HPO_4 and 0.4% glucose, as previously described (37, 75). ChIP was performed as previously described (76), using 1 µl anti-β (RNA polymerase subunit) antibody (BioLegend catalog #663903). ChIP and input samples were analyzed using an ABI 7500 Fast real-time PCR machine. Enrichment of ChIP samples was calculated relative to a control region within *bglB*, which is transcriptionally silent. RNAP Occupancy Units represent background-subtracted fold–enrichment over the control region. Oligonucleotides used for qPCR were JW125 + JW126 (*bglB*), JW10937 + JW10938 (*yoaI*), JW10939 + JW10940 (*amn*), JW10941 + JW10942 (*pstS*), JW10943 + JW10944 (*ugpB*), JW10945 + JW10946 (*mepK*), JW10947 + JW10948 (*gloC*), JW10949 + JW10950 (*evgA*), JW10951 + JW10952 (*evgS*), JW10953 + JW10954 (*pnp* upstream), JW10955 + JW10956 (*pnp* downstream). For statistical analysis of relative changes in RNAP occupancy upstream and downstream of intragenic PhoB sites, we used the mean and standard deviation values for ChIP-qPCR occupancy to generate 1,000 simulations for occupancy at each region tested. We then determined how frequently the ratio of predicted RNAP occupancy downstream:upstream of an intragenic PhoB site was higher in Δ*phoB* cells than wild-type cells. We repeated this simulation 10 times to estimate a *p*-value for each of the three intragenic PhoB sites tested.

Analysis of σ^{70} occupancy around PhoB binding sites

Duplicate ChIP-seq data for σ^{70} from wild-type MG1655, MG1655 Δ*phoB* (DMF84), MG1655 Δ*hns* (AMD565a), and MG1655 Δ*hns* Δ*phoB* (DMF85) were aligned to the *E. coli* MG1655 genome (NC_000913.3) using Rockhopper (version 2.03, default parameters) (77), which also calculated the depth of sequence coverage at all genomic positions on each strand, normalized for total sequence read count. A custom Python script was used to determine the relative sequence read coverage from each σ^{70} ChIP-seq dataset in 400 bp windows centered on each PhoB ChIP-seq peak (coordinates listed in Table 1).

Analysis of σ^{70} occupancy at promoters in wild-type and Δhns strains

Duplicate ChIP-seq data for σ^{70} for wild-type MG1655 or MG1655 Δ*hns* (AMD565a) were aligned to the *E. coli* MG1655 genome (NC_000913.3) using CLC Genomics Workbench (version 8). ChIP-seq peaks were called using a previously described analysis pipeline (8). A custom Python script was used to determine the relative sequence read coverage from each σ^{70} ChIP-seq dataset in 50 bp windows centered on each σ^{70} ChIP-seq peak from MG1655 Δ*hns* (AMD565a).

Determining H-NS occupancy from published ChIP-seq data

H-NS ChIP-seq occupancy was determined from published data (49). Specifically, genome coordinates for σ^{70} ChIP-seq peaks were converted from NCBI genome sequence version U00096.3 to U00096.2 at https://biocyc.org/ECOLI/map-seq-coords-form?chromosome=COLI-K12. H-NS occupancy was determined as the average from four normalized sequence read coverage files available from EBI ArrayExpress, accession number E-MTAB-332.

RNA-seq

MG1655 + pBAD24 or CDS091 (MG1655 Δ*phoB*) + pBAD24 cells were grown at 37° C with aeration in MOPS minimal medium with 0.2 mM K_2PO_4, 0.4% glucose and 100 µg/ml ampicillin to an OD_{600} of 0.5-0.6 Arabinose was added to a final concentration of 0.2% for 7 minutes. Note that addition of arabinose is not expected to impact expression of PhoB-regulated genes. RNA was isolated using a modified hot-phenol method, as previously described (76). Samples were treated with Turbo DNase (Ambion) to remove genomic DNA, ribosomal RNA was removed using the Ribo-Zero rRNA removal kit for Gram-negative bacteria (Epicentre/Illumina), and libraries were prepared with the ScriptSeq Complete kit for bacteria (Epicentre/Illumina) (76). Libraries were sequenced on a HiSeq 2000 (Illumina) by the University at Buffalo Next-Generation Sequencing Core Facility. RNA-seq data were aligned to the *E. coli* MG1655 genome (NC_000913.3) using BWA for Illumina (v0.5.9-r16) (78) on Galaxy (usegalaxy.org) (79). Read counting,

normalization, and differential expression analysis were performed in R using GenomicAlignments (v1.28) *summarizeOverlaps* (80) and DEseq2 (v1.32, betaPrior = FALSE) (81).

PhoB Motif Discovery and Analysis

100 bp sequences surrounding PhoB ChIP-seq peaks were extracted and analyzed using MEME (version 5.1.0, default parameters) (82, 83). The position of the inferred motif relative to ChIP-seq peak centers was analyzed using Centrimo (version 5.1.0, default parameters) (84) through the MEME-ChIP tool (85).

To determine whether the nucleotide content of the PhoB binding site motif contributes to the association of PhoB binding sites with H-NS-bound regions, we first scrambled each PhoB binding site individually using a custom Python script. We then compiled the scrambled sites into a PWM and searched the *E. coli* MG1655 genome (NC_000913.3) for the top 1000 matches to this PWM using FIMO (version 5.1.0, default parameters) (86).

Analysis of PhoB binding site conservation

Binding site conservation analysis was performed as described previously (71). Protein sequences were aligned using Clustal Omega (87) and visualized using MView (88). The genomes analyzed were *Arsenophonus nasoniae* DSM 15247, *Brenneria sp*. EniD312, *Cedecea davisae* DSM 4568, *Citrobacter rodentium* ICC168, *Cronobacter sakazakii* ATCC BAA-894, *Dickeya dadantii* 3937, *Edwardsiella tarda* EIB202, *Enterobacter cloacae* subsp. *cloacae* ATCC 13047, *Erwinia amylovora* ATCC 49946, *Escherichia coli* str. K-12 substr. MG1655, *Hafnia alvei* ATCC 51873, *Klebsiella pneumoniae* KCTC 2242, *Leminorella grimontii* ATCC 33999 = DSM 5078, *Morganella morganii* subsp. *morganii* KT, *Pantoea agglomerans* 299R, *Pectobacterium atrosepticum* SCRI1043, *Photorhabdus asymbiotica*, *Plesiomonas shigelloides* 302-73, *Proteus mirabilis* HI4320, *Providencia stuartii* MRSN 2154, *Pseudomonas aeruginosa* PAO1, *Rahnella sp*. Y9602, *Raoultella*

ornithinolytica B6, *Salmonella enterica* str P125109, *Serratia marcescens* FGI94, *Vibrio cholerae* M66-2, *Xenorhabdus bovienii* SS-2004, *Yersinia pestis* KIM 10, and *Yokenella regensburgei* ATCC 43003.

ACKNOWLEDGEMENTS

We thank the University at Buffalo Next-Generation Sequencing Core Facility and the Wadsworth Center Applied Genomic Technologies Core Facility for DNA sequencing. We thank the Wadsworth Center Tissue Culture and Media Core Facility and Glassware Facility for technical support. This material is based on work supported by the National Science Foundation Graduate Research Fellowship under grant number DGE-1060277 (DMF). DMF was also supported by National Institutes of Health training grant T32AI055429. This work was also supported by a National Institutes of Health Director's New Innovator Award DP2OD007188 (JTW) and National Institutes of Health grants R01GM114812 and R35GM144328 (JTW).

ACCESSION NUMBERS

ChIP-seq data are available at EBI ArrayExpress using accession number E-MTAB-9293. RNA-seq data are available at EBI ArrayExpress using accession number E-MTAB-9591.

REFERENCES

1. Browning DF, Busby SJW. 2016. Local and global regulation of transcription initiation in bacteria. Nature Reviews Microbiology 14:638–650.

2. Galagan J, Lyubetskaya A, Gomes A. 2013. ChIP-Seq and the complexity of bacterial transcriptional regulation. Curr Top Microbiol Immunol 363:43–68.

3. Galagan JE, Minch K, Peterson M, Lyubetskaya A, Azizi E, Sweet L, Gomes A, Rustad T, Dolganov G, Glotova I, Abeel T, Mahwinney C, Kennedy AD, Allard R, Brabant W, Krueger A, Jaini S, Honda B, Yu W-H, Hickey MJ, Zucker J, Garay C, Weiner B, Sisk P, Stolte C, Winkler JK, Van de Peer Y, Iazzetti P, Camacho D, Dreyfuss J, Liu Y, Dorhoi A, Mollenkopf H-J, Drogaris P, Lamontagne J, Zhou Y, Piquenot J, Park ST, Raman S, Kaufmann SHE, Mohney RP, Chelsky D, Moody DB, Sherman DR, Schoolnik GK. 2013. The Mycobacterium tuberculosis regulatory network and hypoxia. Nature 499:178–183.

4. Shimada T, Ishihama A, Busby SJW, Grainger DC. 2008. The Escherichia coli RutR transcription factor binds at targets within genes as well as intergenic regions. Nucleic Acids Res 36:3950–3955.

5. Knapp GS, Lyubetskaya A, Peterson MW, Gomes ALC, Ma Z, Galagan JE, McDonough KA. 2015. Role of intragenic binding of cAMP responsive protein (CRP) in regulation of the succinate dehydrogenase genes Rv0249c-Rv0247c in TB complex mycobacteria. Nucleic Acids Res 43:5377–5393.

6. Ranganathan S, Bai G, Lyubetskaya A, Knapp GS, Peterson MW, Gazdik M, C Gomes AL, Galagan JE, McDonough KA. 2016. Characterization of a cAMP responsive transcription factor, Cmr (Rv1675c), in TB complex mycobacteria reveals overlap with the DosR (DevR) dormancy regulon. Nucleic Acids Res 44:134–151.

7. Brown DR, Barton G, Pan Z, Buck M, Wigneshweraraj S. 2014. Nitrogen stress response and stringent response are coupled in Escherichia coli. Nat Commun 5:4115.

8. Fitzgerald DM, Bonocora RP, Wade JT. 2014. Comprehensive mapping of the Escherichia coli flagellar regulatory network. PLoS Genet 10:e1004649.

9. Bonocora RP, Wade JT. 2015. ChIP-seq for genome-scale analysis of bacterial DNA-binding proteins. Methods Mol Biol 1276:327–340.

10. Haycocks JRJ, Grainger DC. 2016. Unusually Situated Binding Sites for Bacterial Transcription Factors Can Have Hidden Functionality. PLoS ONE 11:e0157016.

11. Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J. 2012. An atlas of Hfq-bound transcripts reveals 3’ UTRs as a genomic reservoir of regulatory small RNAs. EMBO J 31:4005–4019.

12. Smith C, Stringer AM, Mao C, Palumbo MJ, Wade JT. 2016. Mapping the Regulatory Network for Salmonella enterica Serovar Typhimurium Invasion. mBio 7:e01024-16.

13. Irani MH, Orosz L, Adhya S. 1983. A control element within a structural gene: the gal operon of Escherichia coli. Cell 32:783–788.

14. Choy HE, Park SW, Parrack P, Adhya S. 1995. Transcription regulation by inflexibility of promoter DNA in a looped complex. Proc Natl Acad Sci U S A 92:7327–7331.

15. Gielow A, Kücherer C, Kölling R, Messer W. 1988. Transcription in the region of the replication origin, oriC, of Escherichia coli: termination of asnC transcripts. Mol Gen Genet 214:474–481.

16. He B, Zalkin H. 1992. Repression of Escherichia coli purB is by a transcriptional roadblock mechanism. J Bacteriol 174:7121–7127.

17. Belitsky BR, Sonenshein AL. 2011. Roadblock repression of transcription by Bacillus subtilis CodY. J Mol Biol 411:729–743.

18. Belitsky BR, Sonenshein AL. 2013. Genome-wide identification of Bacillus subtilis CodY-binding sites at single-nucleotide resolution. Proc Natl Acad Sci USA 110:7026–7031.

19. Lybecker M, Bilusic I, Raghavan R. 2014. Pervasive transcription: detecting functional RNAs in bacteria. Transcription 5:e944039.

20. Wade JT, Grainger DC. 2014. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. Nat Rev Micro 12:647–653.

21. Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM, Storz G. 2015. Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in Escherichia coli. Journal of Bacteriology 197:18–28.

22. Reitzer LJ, Magasanik B. 1986. Transcription of glnA in E. coli is stimulated by activator bound to sites far from the promoter. Cell 45:785–792.

23. Schleif R. 1992. DNA looping. Annu Rev Biochem 61:199–223.

24. Qian Z, Dimitriadis EK, Edgar R, Eswaramoorthy P, Adhya S. 2012. Galactose repressor mediated intersegmental chromosomal connections in Escherichia coli. Proc Natl Acad Sci USA 109:11336–11341.

25. Qian Z, Trostel A, Lewis DEA, Lee SJ, He X, Stringer AM, Wade JT, Schneider TD, Durfee T, Adhya S. 2016. Genome-Wide Transcriptional Regulation and Chromosome Structural Arrangement by GalR in E. coli. Front Mol Biosci 3:74.

26. Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, Phillips R. 2014. The transcription factor titration effect dictates level of gene expression. Cell 156:1312–1323.

27. Göpel Y, Görke B. 2014. Lies and deception in bacterial gene regulation: the roles of nucleic acid decoys. Mol Microbiol 92:641–647.

28. Mrázek J, Karls AC. 2019. In silico simulations of occurrence of transcription factor binding sites in bacterial genomes. BMC Evol Biol 19:67.

29. Martínez-Hackert E, Stock AM. 1997. Structural relationships in the OmpR family of winged-helix transcription factors. J Mol Biol 269:301–312.

30. Hsieh Y-J, Wanner BL. 2010. Global regulation by the seven-component Pi signaling system. Curr Opin Microbiol 13:198–203.

31. Makino K, Shinagawa H, Amemura M, Kawamoto T, Yamada M, Nakata A. 1989. Signal transduction in the phosphate regulon of Escherichia coli involves phosphotransfer between PhoR and PhoB proteins. J Mol Biol 210:551–559.

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

33. Gardner SG, McCleary WR. 2019. Control of the phoBR Regulon in Escherichia coli. EcoSal Plus 8.

34. Lamarche MG, Wanner BL, Crépin S, Harel J. 2008. The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. FEMS Microbiol Rev 32:461–473.

35. Blus-Kadosh I, Zilka A, Yerushalmi G, Banin E. 2013. The effect of pstS and phoB on quorum sensing and swarming motility in Pseudomonas aeruginosa. PLoS ONE 8:e74444.

36. VanBogelen RA, Olson ER, Wanner BL, Neidhardt FC. 1996. Global analysis of proteins synthesized during phosphorus restriction in Escherichia coli. J Bacteriol 178:4344–4366.

37. Baek JH, Lee SY. 2007. Transcriptome analysis of phosphate starvation response in Escherichia coli. Journal of Microbiology and Biotechnology 17:244–252.

38. Gao R, Stock AM. 2018. Overcoming the Cost of Positive Autoregulation by Accelerating the Response with a Coupled Negative Feedback. Cell Rep 24:3061-3071.e6.

39. Makino K, Shinagawa H, Amemura M, Nakata A. 1986. Nucleotide sequence of the phoB gene, the positive regulatory gene for the phosphate regulon of Escherichia coli K-12. J Mol Biol 190:37–44.

40. Baek JH, Lee SY. 2006. Novel gene members in the Pho regulon of Escherichia coli. FEMS Microbiol Lett 264:104–109.

41. Yoshida Y, Sugiyama S, Oyamada T, Yokoyama K, Makino K. 2012. Novel members of the phosphate regulon in Escherichia coli O157:H7 identified using a whole-genome shotgun approach. Gene 502:27–35.

42. Chekabab SM, Jubelin G, Dozois CM, Harel J. 2014. PhoB activates Escherichia coli O157:H7 virulence factors in response to inorganic phosphate limitation. PLoS ONE 9:e94285.

43. Yang C, Huang T-W, Wen S-Y, Chang C-Y, Tsai S-F, Wu W-F, Chang C-H. 2012. Genome-wide PhoB binding and gene expression profiles reveal the hierarchical gene regulatory network of phosphate starvation in Escherichia coli. PloS one 7:e47314.

44. Reppas NB, Wade JT, Church GM, Struhl K. 2006. The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. Mol Cell 24:747–757.

45. Singh SS, Singh N, Bonocora RP, Fitzgerald DM, Wade JT, Grainger DC. 2014. Widespread suppression of intragenic transcription initiation by H-NS. Genes and Development 28:214–219.

46. Blanco AG, Canals A, Bernués J, Solà M, Coll M. 2011. The structure of a transcription activation subcomplex reveals how σ(70) is recruited to PhoB promoters. The EMBO journal 30:3776–85.

47. Canals A, Blanco AG, Coll M. 2012. σ70 and PhoB activator: Getting a better grip. Transcription https://doi.org/10.4161/trns.20444.

48. Tung C-S, McMahon BH. 2012. A structural model of the E. coli PhoB dimer in the transcription initiation complex. BMC structural biology 12:3.

49. Kahramanoglou C, Seshasayee ASN, Prieto AI, Ibberson D, Schmidt S, Zimmermann J, Benes V, Fraser GM, Luscombe NM. 2011. Direct and indirect effects of H-NS and Fis on global gene expression control in Escherichia coli. Nucleic acids research 39:2073–91.

50. Grainger DC. 2016. Structure and function of bacterial H-NS protein. Biochem Soc Trans 44:1561–1569.

51. Lamberte LE, Baniulyte G, Singh SS, Stringer AM, Bonocora RP, Stracy M, Kapanidis AN, Wade JT, Grainger DC. 2017. Horizontally acquired AT-rich genes in Escherichia coli cause toxicity by sequestering RNA polymerase. Nature Microbiology 2:16249–16249.

52. Zhang Z, Gerstein M. 2003. Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements. J Biol 2:11.

53. Suziedelienė E, Suziedélis K, Garbenciūtė V, Normark S. 1999. The acid-inducible asr gene in Escherichia coli: transcriptional control by the phoBR operon. J Bacteriol 181:2084–2093.

54. Han JS, Park JY, Lee YS, Thöny B, Hwang DS. 1999. PhoB-dependent transcriptional activation of the iciA gene during starvation for phosphate in Escherichia coli. Mol Gen Genet 262:448–452.

55. Harris RM, Webb DC, Howitt SM, Cox GB. 2001. Characterization of PitA and PitB from Escherichia coli. J Bacteriol 183:5008–5014.

56. Murray EL, Conway T. 2005. Multiple regulators control expression of the Entner-Doudoroff aldolase (Eda) of Escherichia coli. J Bacteriol 187:991–1000.

57. Yoshida Y, Sugiyama S, Oyamada T, Yokoyama K, Kim S-K, Makino K. 2011. Identification of PhoB binding sites of the yibD and ytfK promoter regions in Escherichia coli. J Microbiol 49:285–289.

58. Jia T, Wu P, Liu B, Liu M, Mu H, Liu D, Huang M, Li L, Wei Y, Wang L, Yang Q, Liu Y, Yang B, Huang D, Yang L, Liu B. 2023. The phosphate-induced small RNA EsrL promotes E. coli virulence, biofilm formation, and intestinal colonization. Sci Signal 16:eabm0488.

59. Lubin EA, Henry JT, Fiebig A, Crosson S, Laub MT. 2016. Identification of the PhoB Regulon and Role of PhoU in the Phosphate Starvation Response of Caulobacter crescentus. J Bacteriol 198:187–200.

60. Izutsu K, Wada A, Wada C. 2001. Expression of ribosome modulation factor (RMF) in Escherichia coli requires ppGpp. Genes Cells 6:665–676.

61. Sanchez-Vazquez P, Dewey CN, Kitten N, Ross W, Gourse RL. 2019. Genome-wide effects on Escherichia coli transcription from ppGpp binding to its two sites on RNA polymerase. Proc Natl Acad Sci USA 116:8310–8319.

62. Shimada T, Yoshida H, Ishihama A. 2013. Involvement of cyclic AMP receptor protein in regulation of the rmf gene encoding the ribosome modulation factor in Escherichia coli. J Bacteriol 195:2212–2219.

63. Yoshida H, Shimada T, Ishihama A. 2018. Coordinated Hibernation of Transcriptional and Translational Apparatus during Growth Transition of Escherichia coli to Stationary Phase. mSystems 3.

64. Moen B, Janbu AO, Langsrud S, Langsrud O, Hobman JL, Constantinidou C, Kohler A, Rudi K. 2009. Global responses of Escherichia coli to adverse conditions determined by microarrays and FT-IR spectroscopy. Can J Microbiol 55:714–728.

65. Makino K, Amemura M, Kim SK, Nakata A, 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.

66. Grainger DC. 2016. The unexpected complexity of bacterial genomes. Microbiology (Reading, Engl) 162:1167–1172.

67. Wade JT, Castro Roa D, Grainger DC, Hurd D, Busby SJW, Struhl K, Nudler E. 2006. Extensive functional overlap between sigma factors in Escherichia coli. Nature structural & molecular biology 13:806–14.

68. Bonocora RP, Fitzgerald DM, Stringer AM, Wade JT. 2013. Non-canonical protein-DNA interactions identified by ChIP are not artifacts. BMC Genomics 14:254.

69. Bonocora RP, Smith C, Lapierre P, Wade JT. 2015. Genome-Scale Mapping of Escherichia coli σ54 Reveals Widespread, Conserved Intragenic Binding. PLoS Genet 11:e1005552.

70. Wong GT, Bonocora RP, Schep AN, Beeler SM, Lee Fong AJ, Shull LM, Batachari LE, Dillon M, Evans C, Becker CJ, Bush EC, Hardin J, Wade JT, Stoebel DM. 2017. Genome-Wide Transcriptional Response to Varying RpoS Levels in Escherichia coli K-12. J Bacteriol 199.

71. Fitzgerald DM, Smith C, Lapierre P, Wade JT. 2018. The evolutionary impact of intragenic FliA promoters in proteobacteria. Mol Microbiol 108:361–378.

72. Stringer AM, Singh N, Yermakova A, Petrone BL, Amarasinghe JJ, Reyes-Diaz L, Mantis NJ, Wade JT. 2012. FRUIT, a scar-free system for targeted chromosomal mutagenesis, epitope tagging, and promoter replacement in Escherichia coli and Salmonella enterica. PloS one 7:e44841.

73. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko K a, Tomita M, Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular systems biology 2:2006.0008.

74. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci USA 97:6640–6645.

75. Neidhardt FC, Bloch PL, Smith DF. 1974. Culture Medium for Enterobacteria Culture Medium for Enterobacteria 119:736–747.

76. Stringer AM, Currenti SA, Bonocora RP, Baranowski C, Petrone BL, Palumbo MJ, Reilly A a E, Zhang Z, Erill I, Wade JT. 2014. Genome-Scale Analyses of Escherichia coli and Salmonella enterica AraC Reveal Noncanonical Targets and an Expanded Core Regulon. Journal of bacteriology 196:660–71.

77. McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco C a, Vanderpool CK, Tjaden B. 2013. Computational analysis of bacterial RNA-Seq data. Nucleic acids research 41:e140.

78. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

79. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research 46:W537–W544.

80. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for Computing and Annotating Genomic Ranges. PLOS Computational Biology 9:e1003118.

81. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15:550–550.

82. Bailey TL, Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2:28–36.

83. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. Nucleic acids research 37:W202-8.

84. Bailey TL, Machanick P. 2012. Inferring direct DNA binding from ChIP-seq. Nucleic Acids Res 40:e128.

85. Machanick P, Bailey TL. 2011. MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics 27:1696–1697.

86. Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. Bioinformatics 27:1017–1018.

87. Sievers F, Higgins DG. 2014. Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol 1079:105–116.

88. Brown NP, Leroy C, Sander C. 1998. MView: a web-compatible database search or multiple alignment viewer. Bioinformatics 14:380–381.

89. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of Escherichia coli K-12. Science (New York, NY) 277:1453–1474.

90. Guzman LM, Belin D, Carson MJ, Beckwith J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. Journal of bacteriology 177:4121–30.

FIGURE LEGENDS

Figure 1. Partially reduced activity of the C-terminally FLAG$_3$-tagged PhoB. Quantitative RT-PCR was used to measure levels of the *pstS* RNA relative to the *minD* RNA control in wild-type MG1655 + pBAD24 ("Wild-type"), MG1655 Δ*phoB* (CDS091) + pBAD24 ("Δ*phoB*"), or MG1655 *phoB*-FLAG$_3$ (DMF34) + pBAD24 ("PhoB-FLAG$_3$"), for cells grown under low phosphate conditions. Values represent the average of three independent biological replicates; error bars represent +/- one standard deviation.

Figure 2. ChIP-seq identifies PhoB binding sites. (A) ChIP-seq data for (i) an untagged control under low phosphate conditions, (ii) PhoB-FLAG$_3$ under low phosphate conditions, and (iii) PhoB-FLAG$_3$ under high phosphate conditions. Three genomic regions are shown, with one dataset from two independent biological replicates. Values on the *x*-axis represent genome position. Values on the *y*-axis represent normalized sequence read coverage, with positive values indicating sequence reads mapping to the forward strand, and negative values indicating sequence reads mapping to the reverse strand. *y*-axis scales differ between the three genomic regions but are matched for the three datasets for any given genomic region. **(B)** Significantly enriched DNA sequence motif derived from 100 bp regions surrounding each ChIP-seq peak. The number of sites contributing to the motif and the *E*-value determined by MEME are indicated. **(C)** Analysis of the position of inferred PhoB binding sites relative to the position of ChIP-seq peak centers. For each of the binding sites contributing to the motif determined by MEME (see panel (C)), we determined the position of the binding site relative to the associated ChIP-seq peak center. The *x*-axis indicates position relative to ChIP-seq peak centers. The *y*-axis indicates the number of binding sites that cover any given position. **(D)** Pie-chart showing the genome context of PhoB binding sites identified by ChIP-seq. Sites designated as "Intragenic & Upstream" are intragenic but <200 bp upstream of an annotated gene start.

Figure 3. Comparison of ChIP-seq and ChIP-chip datasets. (A) Significantly enriched DNA sequence motif derived from 100 bp regions surrounding each ChIP-seq peak for regions shared between the ChIP-seq dataset and a published ChIP-chip dataset (43). The number of sites contributing to the motif, and the *E*-value determined by MEME are indicated. **(A)** Significantly enriched DNA sequence motif derived from 100 bp regions surrounding each ChIP-seq peak for regions unique to the ChIP-seq dataset, i.e. not found in the published ChIP-chip dataset (43). The number of sites contributing to the motif, and the *E*-value determined by MEME are indicated.

Figure 4. RNA-seq analysis of wild-type and Δ*phoB E. coli*. Scatter-plot showing relative RNA levels for all genes in wild-type (MG1655 + pBAD24) or Δ*phoB* (CDS091 + pBAD24) cells. Each datapoint corresponds to a gene. Triangle datapoints represent genes previously reported to be in the *pho* regulon, with red fill indicating that the transcript has an upstream PhoB site identified by ChIP-seq, and gray fill indicating no upstream site. Red circle datapoints represent genes not previously reported to be in the *pho* regulon but with upstream PhoB sites identified by ChIP-seq. Blue circle datapoints represent genes with internal PhoB sites identified by ChIP-seq. All other genes are represented by gray circle datapoints.

Figure 5. Differences in RNAP (β) occupancy in genes that are potential members of the *pho* regulon. RNAP (β) occupancy measured by ChIP-qPCR (see Methods for details of how occupancy units are calculated) in wild-type MG1655 (dark bars) and MG1655 Δ*phoB* (CDS091; light bars) for regions within genes that are potential members of the *pho* regulon. Schematics on the left show genes with upstream or internal PhoB sites (red vertical lines). The horizontal bars indicate the positions of PCR amplicons used in ChIP-qPCR; black bars indicate amplicons within genes that have upstream PhoB sites, green bars indicate amplicons upstream of intragenic PhoB sites, and blue bars indicate amplicons downstream of intragenic PhoB sites. Values represent the average of three independent biological replicates; error bars represent one standard deviation.

Figure 6. Differences in σ^{70} occupancy around PhoB binding sites between wild-type and Δ*phoB* cells. The scatter-plot shows normalized σ^{70} occupancy in wild-type MG1655 and MG1655 Δ*phoB* (DMF84) for the 400 bp regions surrounding PhoB binding sites identified by ChIP-seq. Each datapoint represents a PhoB binding site. Intergenic binding sites are indicated by red datapoints; intragenic binding sites by blue datapoints. Genes associated with PhoB binding sites are labeled with the gene name in cases where σ^{70} occupancy differs >2-fold between wild-type and Δ*phoB* cells. Values represent the average of two independent biological replicates; error bars represent +/- one standard deviation.

Figure 7. H-NS suppresses transcription from many promoters. (A) The scatter-plot shows normalized σ^{70} occupancy in wild-type MG1655 and MG1655 Δ*hns* (AMD565a) for the 400 bp regions surrounding PhoB binding sites identified by ChIP-seq. Each datapoint represents a PhoB binding site. The color of each datapoint indicates the level of H-NS occupancy at the corresponding site (49). Intragenic PhoB sites are represented by crosses; intergenic PhoB sites are represented by circles. **(B)** The scatter-plot shows normalized σ^{70} occupancy in wild-type MG1655 and MG1655 Δ*hns* (AMD565a) for all σ^{70} binding sites identified by ChIP-seq from MG1655 Δ*hns* (AMD565a) cells. The color of each datapoint indicates the level of H-NS occupancy at the corresponding site (49). Intragenic PhoB sites are represented by crosses; intergenic PhoB sites are represented by circles. For both (A) and (B), values represent the average of two independent biological replicates; error bars represent +/- one standard deviation.

Figure 8. H-NS does not suppress PhoB-dependent effects on recruitment of initiating RNA polymerase. The scatter-plot shows normalized σ^{70} occupancy in wild-type MG1655 Δ*hns* (AMD565a) and MG1655 Δ*hns* Δ*phoB* (DMF85) for the 400 bp regions surrounding PhoB binding sites identified by ChIP-seq. Each datapoint represents a PhoB binding site. The color of each datapoint indicates the level of H-NS occupancy at the corresponding site (49). Intragenic PhoB sites are represented by crosses; intergenic PhoB sites are represented by circles. Genes associated with PhoB binding sites are indicated in cases where σ^{70} occupancy differs >2-fold

between MG1655 Δ*hns* (AMD565a) and MG1655 Δ*hns* Δ*phoB* (DMF85) cells. Values represent the average of two independent biological replicates; error bars represent +/- one standard deviation.

Figure 9. Conservation of PhoB binding sites across γ-proteobacterial species. Heat map showing conservation of PhoB binding sites across selected γ-proteobacterial species. Columns represent PhoB binding sites from *E. coli*, divided into known *pho* regulon binding sites, intergenic sites, and intragenic sites. The associated genes are indicated above each column. Rows represent species for different γ-proteobacterial genera, as indicated to the left of each row. The color of each square indicates the predicted strength of the best-scoring putative PhoB binding site in a region that is homologous to the corresponding region in *E. coli*. Binding site strength was predicted using a position weight matrix derived from the *E. coli* PhoB binding site motif (Figure 2B). The color scale is shown below the heat map, with yellow indicating stronger predicted binding site strength, and blue indicating weaker predicted binding site strength. White indicates the absence of a homologous region in the indicated species.

Table 1. List of PhoB-bound regions identified by ChIP-seq.

Genome Coordinate[1]	H-NS-bound?[2]	ChIP Score[3]	Associated Genes[4]	Binding Site Sequence[5]	Expression wt[6]	Expression Δ*phoB*[6,7]
14763		1	(*dnaJ*)	N/A	(8326)	(4689)
22367		2	*ileS*	TGTAATCAAACCGAAATA	21580	24295
201079		1	(*lpxD*)	GGTCACATTACGTTCATG	(9261)	(11591)
260277	x	14	*phoE*/*proB*	TGTAATAAAAGCGTAAAC	68155/6306	69*/4330
262258		1	(*proA*)	TGTAATACGGTTGAAACG	(9578)	(6248)
332492	x	3	(*yahA*)	TGTAACAGAAATATCACA	(2246)	(1084*)
401676	x	20	*phoA*	TGTCATAAAGTTGTCACG	310112	1548*
417093		11	*sbcD*/*phoB*	TTTCATAAATCTGTCATA	274/32731	271/34*
417245		5	(*phoB*)	ATTCACAGCACTGTCATA	(32731)	(34*)
569907	x	6	(*ybcK*)	TGTCACATCGATGTAATC	(33)	(24)
595521		3	*cusR*/*cusC*	TGTCATTTTTCTGTCACC	391/208	239/50*
922786		2	*cspD*/*clpS*	TGTCACATTCCTGTCAAT	2330/18217	1603/23285
983795		2	(*gloC*)	TGTCAGGCCGCTGTCATC	(4405)	(7533*)
1015596		7	*rmf*	ATTCACGCCACTGTCATA	3816	5371
1065537	x	2	*agp*	AGTCATATTTCTGTCACA	976	1983*
1084845		11	*phoH*	TGTCATCACTCTGTCATC	37076	13375*
1096396	x	1	(*ycdU*)	TGTCACAAAAGAATCACT	(53)	(64)
1117391		1	(*yceA*)	GATAAAAAAATCGTCATG	(560)	(456)
1371360	x	5	(*ycjM*)	GGTCACATTTATTTCATA	(11)	(14)
1447326		2	*feaR*/*feaB*	TTTCACAGAGCGAAAACG	111/786	379*/1231
1527862	x	1	*rhsE*	TATCAGAAAAATGTCATG	N.D.	N.D.
1577158	x	4	(*yddB*)	CGGCACAAAACTGTCATA	(94)	(99)
1581949	x	2	(*ydeN*)	TGTCAAAAATCAGTAATG	(120)	(119)
1861798		5	(*yeaC*)	N/A	(1355)	(1480)
1874218		1	*yoaI*/*yeaL*	TGTCATCAAACTGCCATT	12266/36	28*/67
1948542		1	(*nudB*)	CGTCACGCCGCTGCAACA	(3654)	(3763)
1978157		1	(*flhD*)	TGACATCAACTTGTCATA	(71)	(57)
2055017		1	*amn*	TTTCACATTTCTGTGACA	42299	12429*
2137862	x	5	*wza*/*yegH*	TGTCACAATTCGATCATG	10/2257	15/2498
2240543		6	*mglB*	TGTAACCCGTATGTAACA	1076	772
2438961		8	(*yfcJ*)	TGTCACGATACTGTCATT	(163)	(187)
2484480	x	2	(*evgS*)	AGTAACAACCGTGTCACA	(1514)	(3343*)
2743442		3	(*yfiN*) *yfiB*	TGACACAAATCTTTAATC	(575) 1293	(656) 1629
2799166		1	(*alaE*)	AGTCACGCTTGCGTCATG	(83)	(58)
2819248		1	*csrA*	TGTAATGTGTTTGTCATT	6539	7460
2976292		4	*omrA*	GGTCATCAATCTGTAACA	1	3
3031605		3	(*uacT*)	CGTCACATTATTGCAATG	(5)	(4)
3081080		2	(*tktA*)	AGAAATACCGTTGTCATC	(26875)	(27000)
3198259		1	(*glnE*)	TGTCATCTTCCTGCAACG	(11291)	(9028)
3243474		3	(*uxaC*) *uxaA*	TGTCATACACCCGTCACG	(608) 243	(465) 228

3309385		1	(*pnp*)	CTTCACAGTACCGTCATC	(33309)	(53983*)
3362763	x	4	(*yhcA*) *yhcD*	GGTAATAAATATGTCACT	(38) 425	(54) 415
3458489		1	(*gspE*)	N/A	(69)	(51)
3573580		5	*glgB*	GGTCAAAAAAATGTCACA	21445	17405
3592453		35	*ugpB*	TGTCATCTTTCTGACACC	54461	2830*
3600942		2	*rpoH*	TTTCATCTCTATGTCACA	7346	4943
3648534		2	(*arsR*)	GGTAACAGAAATGACATA	(61)	(66)
3672318	x	2	*yhjB*/*yhjC*	TTTCACAATGTTGTCATG	252/397	218/408
3682037		1	*yhjJ*	GAACATGAAAATGTCACG	1871	2721
3709177		1	(*eptB*)	GGTCACCGAGTTGTCATA	(585)	(1126*)
3728905		1	(*xylB*)	AGTAATCTTTCGGTAATA	(1160)	(836)
3731396	x	1	(*xylF*)	N/A	(46)	(31)
3767949	x	2	(*rhsJ*)	TTACATACAAATGTAATA	(N.D.)	(N.D.)
3776578		6	*yibT*/*yibL*	TGTAATAGTTCTGTAACG	5264/1129	2601*/2135*
3911600		38	*pstS*	TGTCATAAAACTGTCATA	239097	310*
3937206		3	*rbsK*	TGTCACCATCAGGTCATA	1452	1701
3990899		1	*hemC*/*cyaA*	TTTCACGCCGTTGTAATA	3823/16054	5169/18141
4113746	x	1	(*fpr*)	CGTAAATGTTTCGTCATC	(4069)	(4643)
4128495		2	*metJ*/*metB*	N/A	543/1432	524/1263
4142314		1	*frwA*/*frwC*	TGTAATGTAACCGTCAAT	72/35	60/36
4227508		1	(*metH*)	ATTCACAAATCTGTCACT	(18026)	(25993)
4245063		1	(*malF*)	TGTCATTAAAAAGAAACA	(486)	(853*)
4325211		2	*phnC*	ATTAACCAAATCGTCACA	32470	13*
4459250		1	(*pmbA*)	GGTAACGATATTGAAACA	(11462)	(11004)
4522523	x	1	(*yjhG*)	N/A	(478)	(444)

[1] Position of the center of the PhoB-bound region in the *E. coli* MG1655 genome (NC_000913.3). Positions in red text indicate those that overlap with regions found by ChIP-chip by (43).
[2] An "x" indicates that the region was previously reported to be bound by H-NS by (49).
[3] A measure of relative enrichment from ChIP-seq data.
[4] Genes in parentheses indicate an intragenic binding site. Downstream genes are listed if the PhoB ChIP-seq peak center is in an intergenic region upstream of that gene, and/or if the peak center is <200 bp upstream of the gene start. Previously described *pho* regulon members are underlined.
[5] Binding sites identified by MEME. Red sequences match binding sites described by (43). "N/A" indicates ChIP-seq peaks for which no binding site sequence was found by MEME.
[6] Relative RNA levels in MG1655 + pBAD24 (wt) or MG1655 Δ*phoB* (CDS091) + pBAD24 (Δ*phoB*) for genes indicated in the third column. "N/D" indicates genes for which RNA levels were not determined.
[7] Asterisks indicate expression differences between the wild-type and Δ*phoB* strains that were determined to be statistically significant ($q < 0.01$).

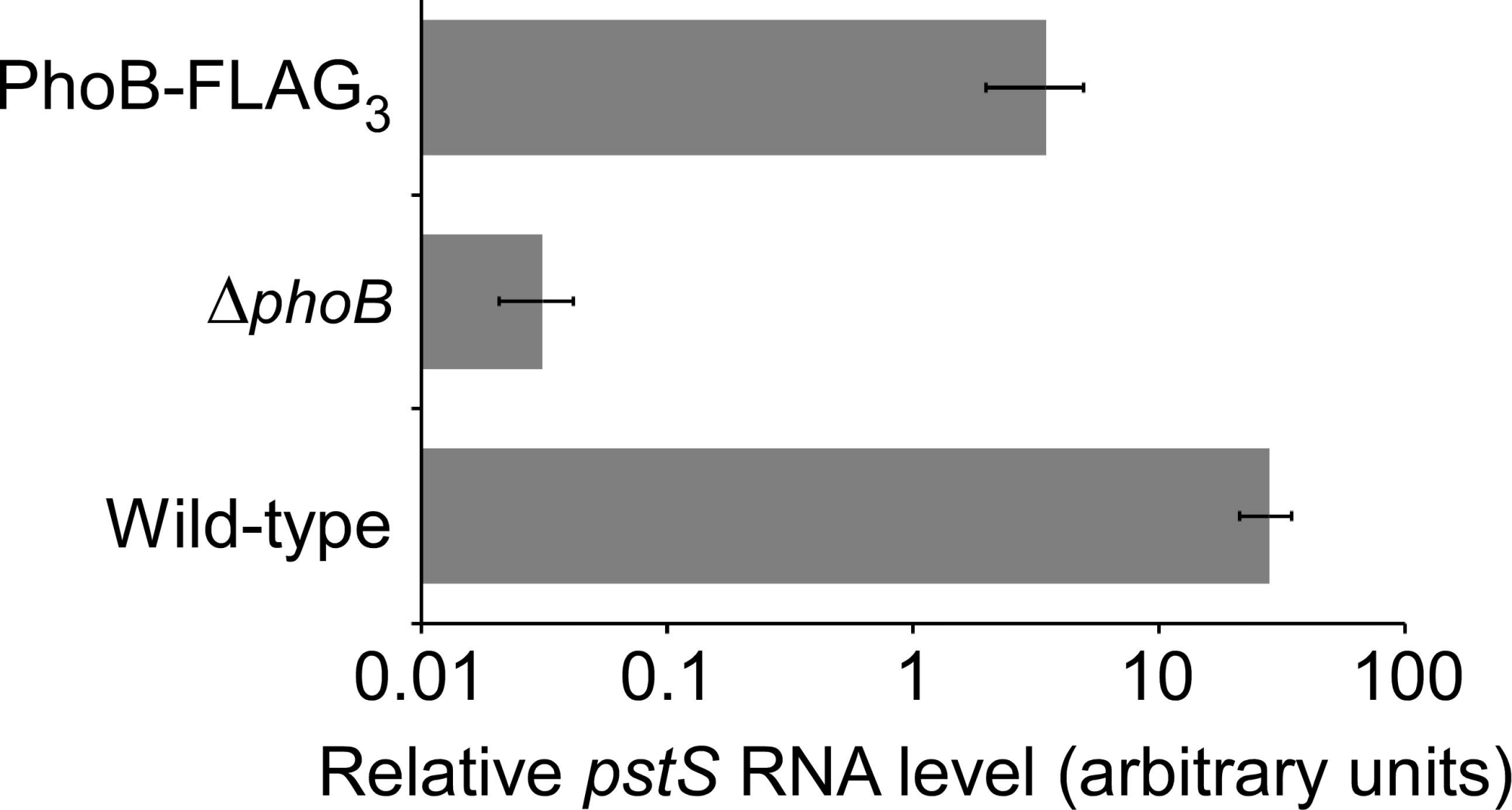
Table 2. List of strains and plasmid used in this study.

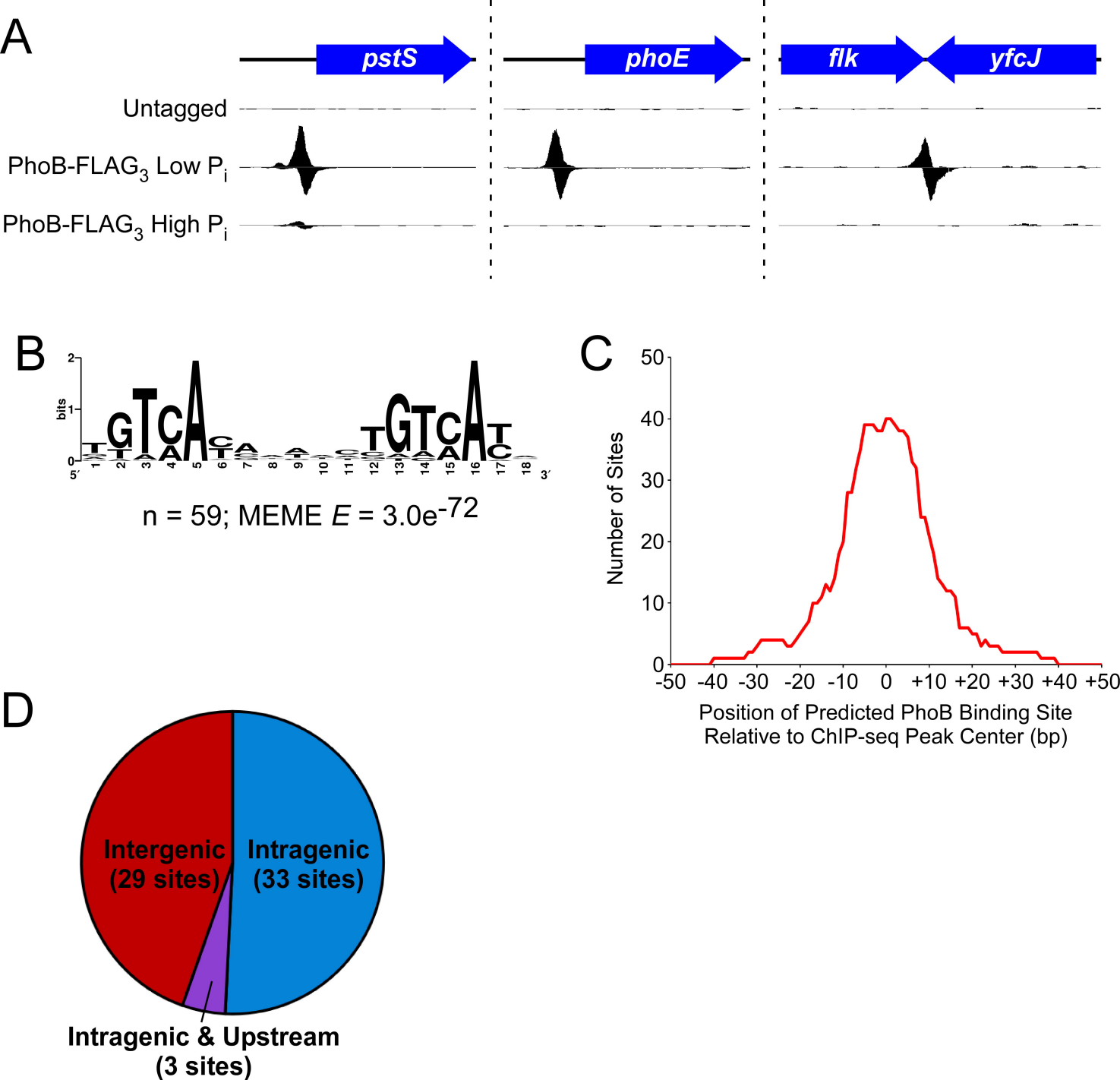
Strain	**Description**	**Source**
MG1655	*Escherichia coli* MG1655	(89)
DMF34	MG1655 *phoB*-FLAG$_3$	This study
DMF84	MG1655 Δ*phoB*	This study
AMD565a	MG1655 Δ*hns*	This study
DMF85	MG1655 Δ*hns* Δ*phoB*	This study
CDS091	MG1655 Δ*phoB*	This study
Plasmid	**Description**	**Source**
pBAD24	Empty vector for arabinose-inducible expression	(90)

SUPPLEMENTARY TABLES

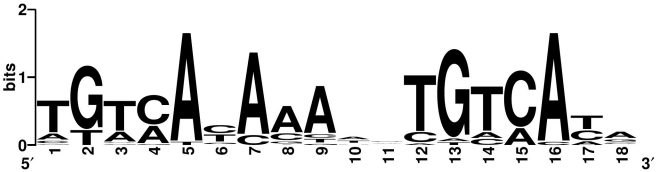
Table S1. RNA-seq analysis.

Table S2. List of oligonucleotides used in this study.


PhoB-FLAG3
ΔphoB
Wild-type
0.01
0.1
1
10
100
Relative pstS RNA level (arbitrary units)


A
pstS
phoE
flk
yfcJ
Untagged
PhoB-FLAG3 Low Pi
PhoB-FLAG3 High Pi
B
bits
n = 59; MEME E = 3.0e-72
C
Number of Sites
Position of Predicted PhoB Binding Site
Relative to ChIP-seq Peak Center (bp)
D
Intergenic
(29 sites)
Intragenic
(33 sites)
Intragenic & Upstream
(3 sites)

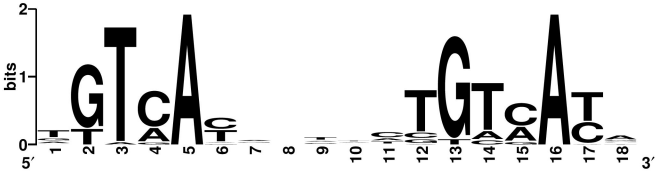
A



n = 25; MEME E = 4.5e^{-27}

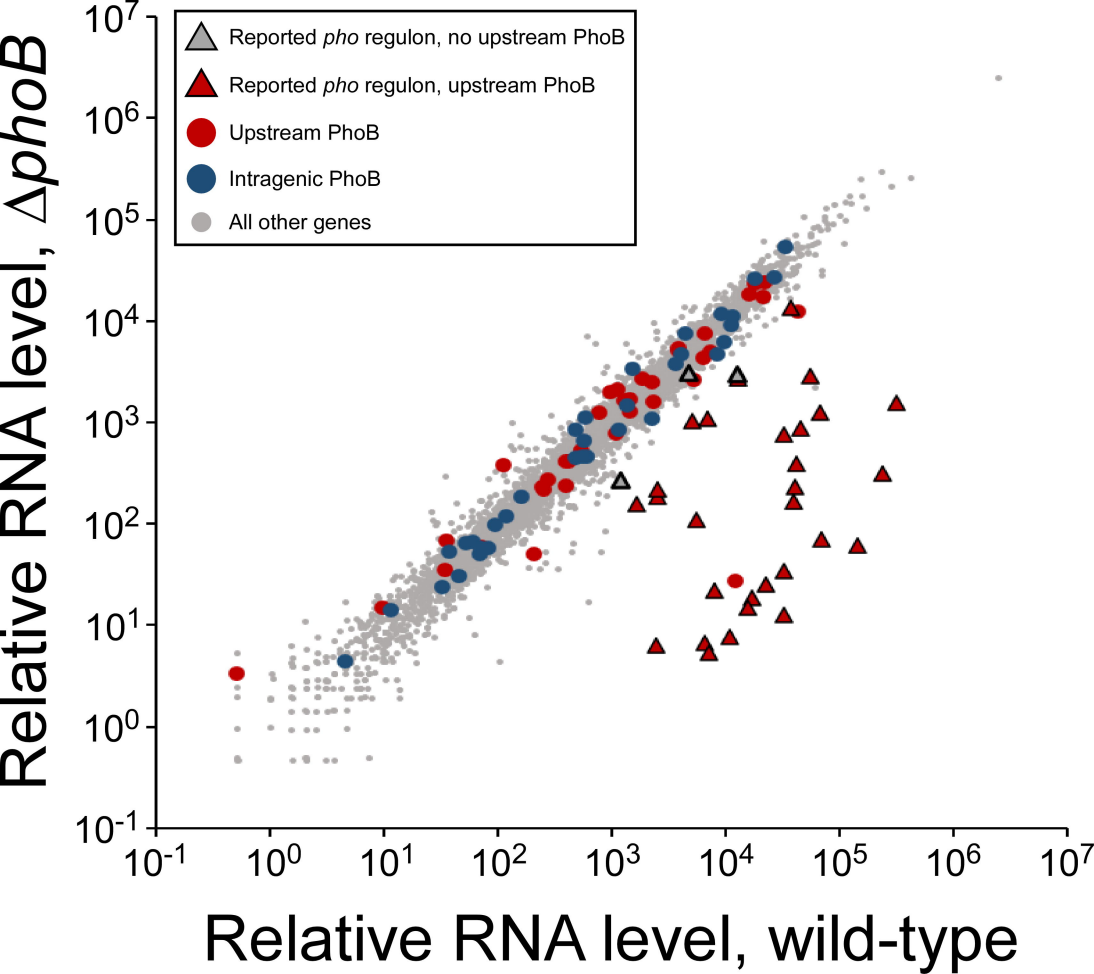
PhoB-bound regions shared with Yang *et al*. (2012)

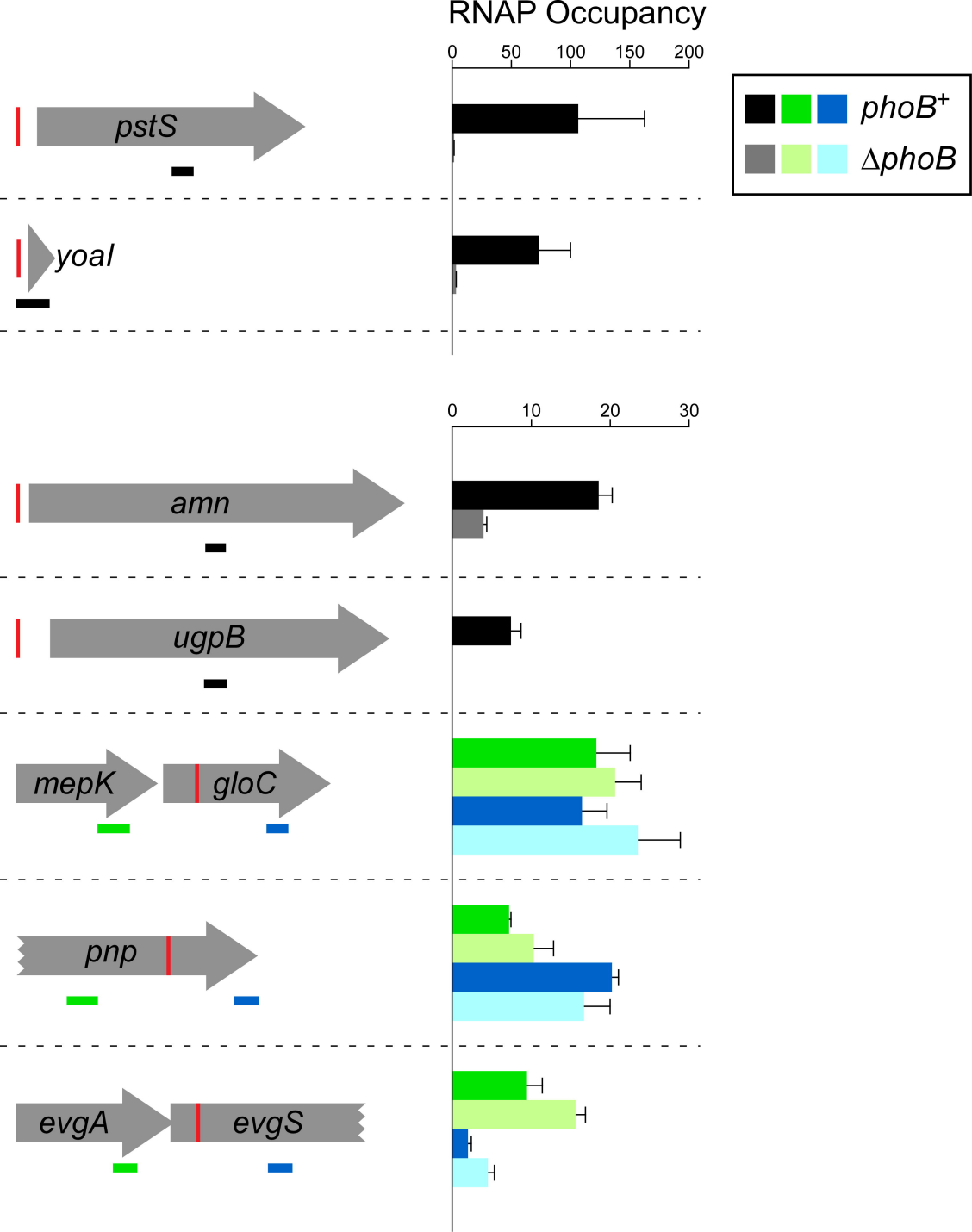
B

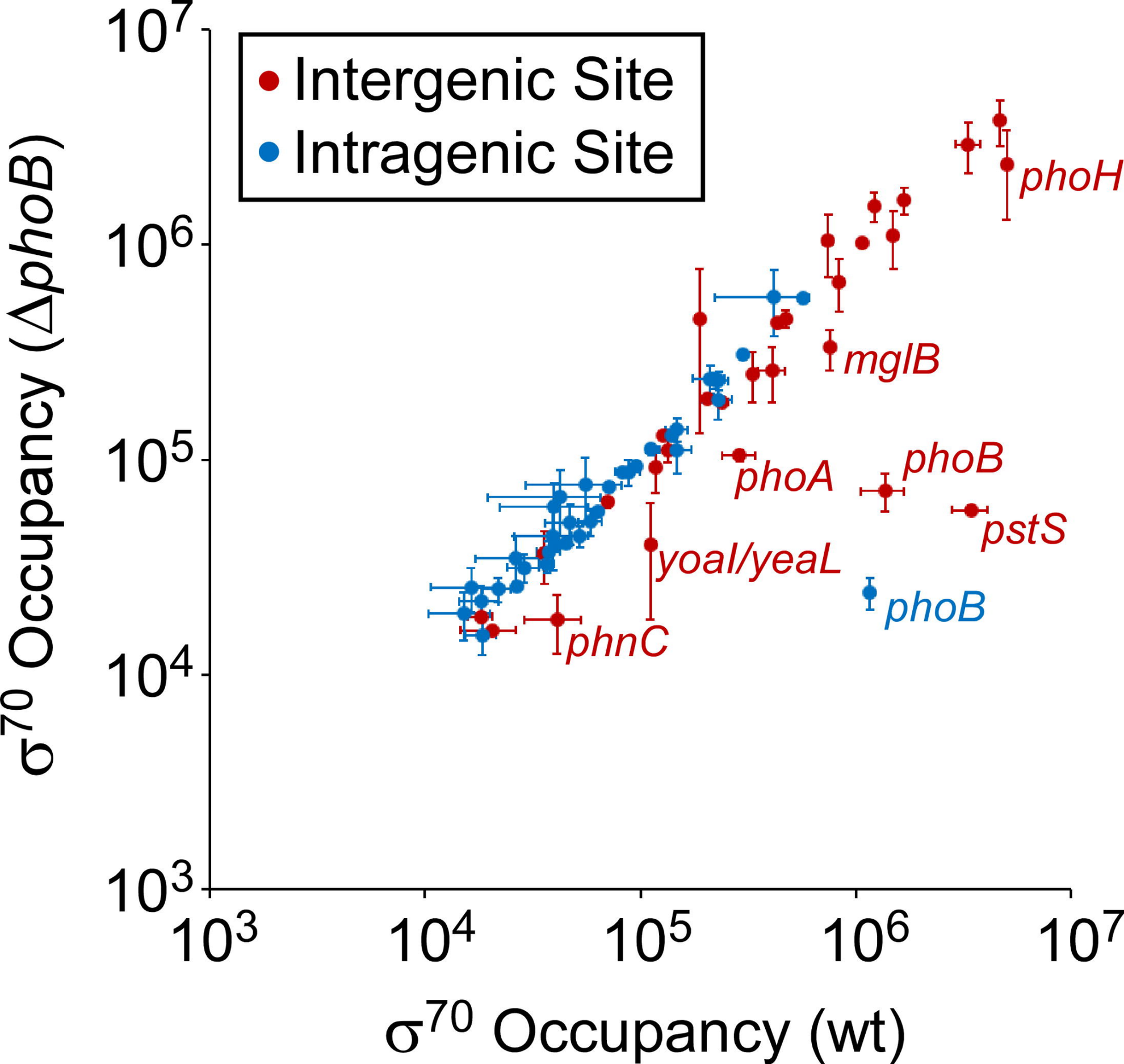


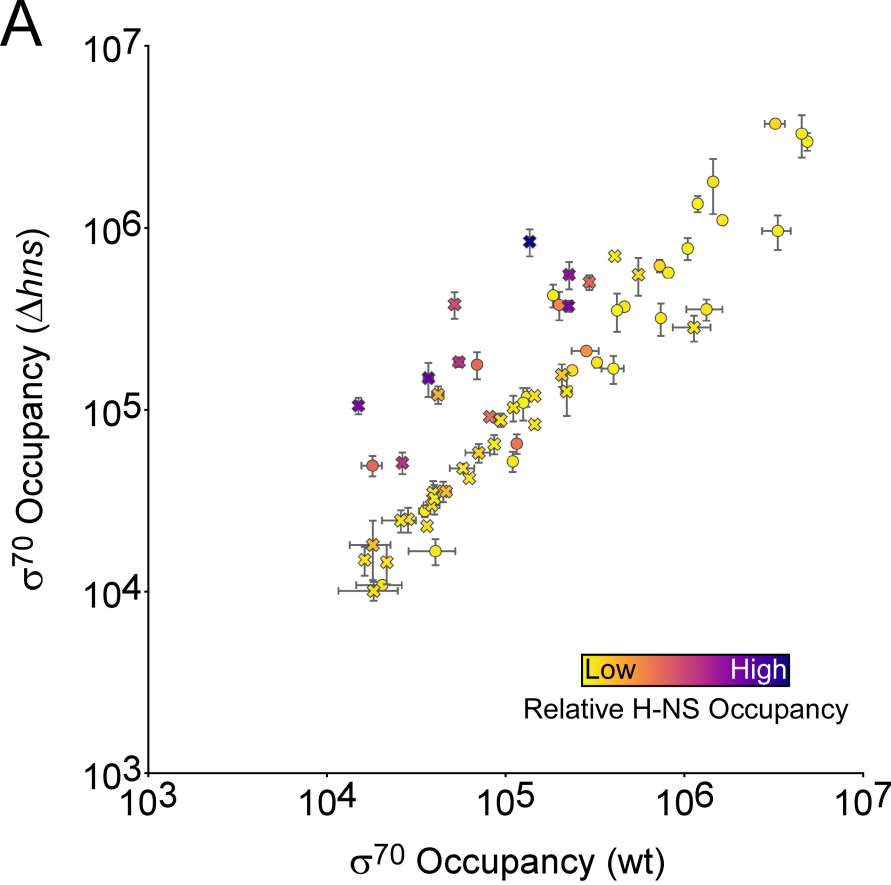
n = 34; MEME E = 2.7e^{-27}

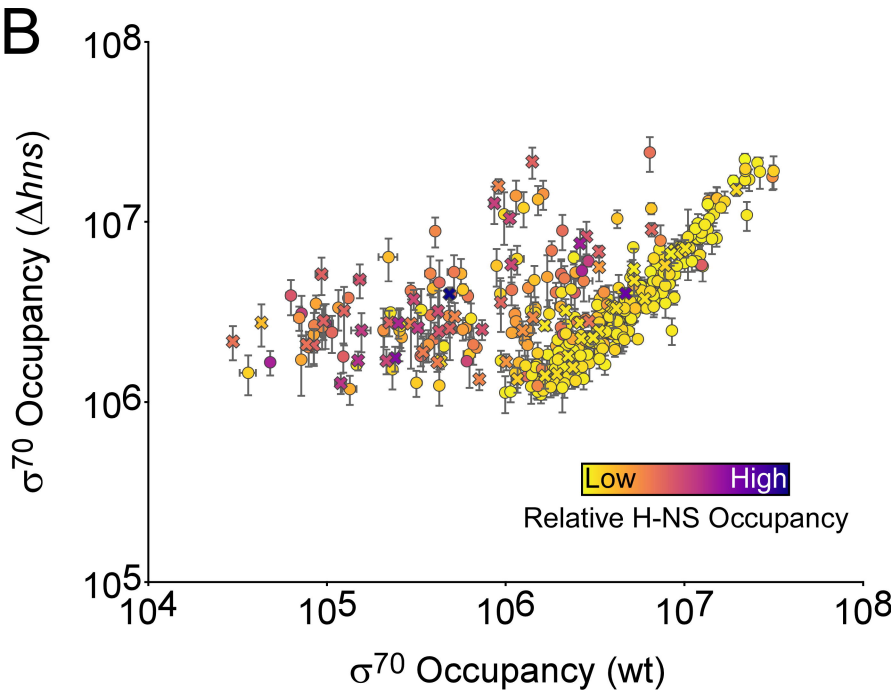
PhoB-bound regions unique to this study

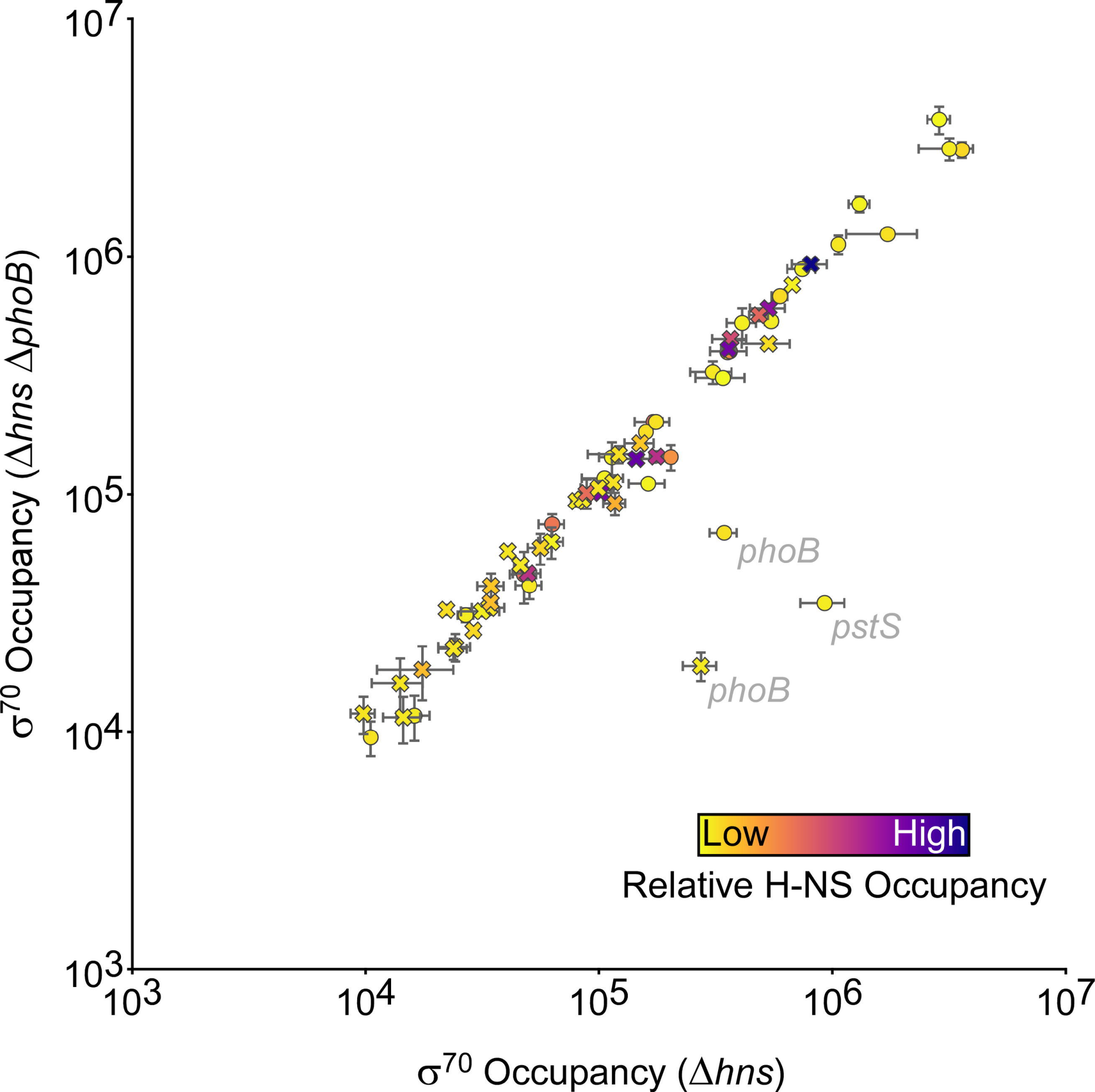

Reported *pho* regulon, no upstream PhoB
Reported *pho* regulon, upstream PhoB
Upstream PhoB
Intragenic PhoB
All other genes
Relative RNA level, Δ*phoB*
Relative RNA level, wild-type
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
10^{5}
10^{6}
10^{7}

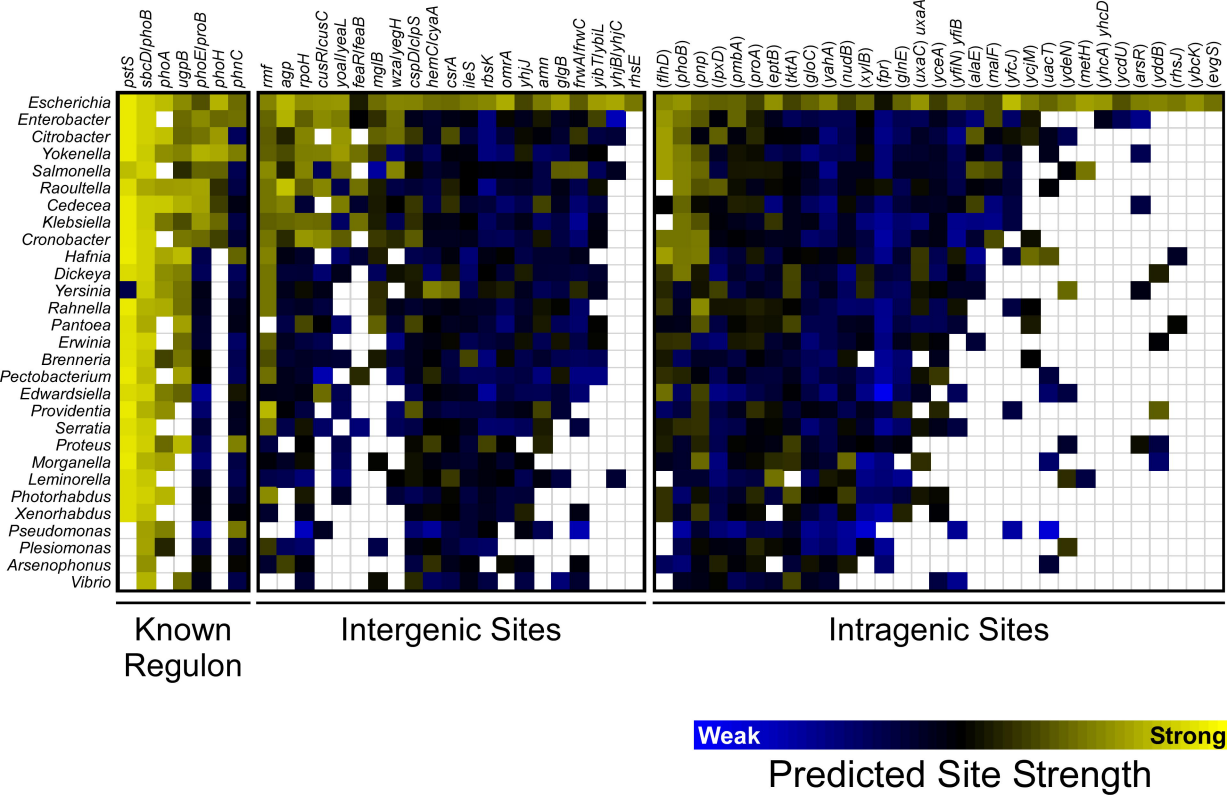

RNAP Occupancy
0
50
100
150
200
phoB+
ΔphoB
pstS
yoaI
0
10
20
30
amn
ugpB
mepK
gloC
pnp
evgA
evgS


Intergenic Site
Intragenic Site
phoH
mglB
phoA
phoB
pstS
yoal/yeaL
phoB
phnC
10^7
10^6
10^5
10^4
10^3
10^3
10^4
10^5
10^6
10^7
σ^{70} Occupancy (Δ*phoB*)
σ^{70} Occupancy (wt)


A
10^7
10^6
10^5
10^4
10^3
σ^{70} Occupancy (Δhns)
Low
High
Relative H-NS Occupancy
10^3
10^4
10^5
10^6
10^7
σ^{70} Occupancy (wt)


B
10^8
10^7
10^6
10^5
σ^{70} Occupancy (Δhns)
Low
High
Relative H-NS Occupancy
10^4
10^5
10^6
10^7
10^8
σ^{70} Occupancy (wt)


10^7
10^6
10^5
10^4
10^3
σ^{70} Occupancy (Δ*hns* Δ*phoB*)
phoB
pstS
phoB
Low
High
Relative H-NS Occupancy
10^3
10^4
10^5
10^6
10^7
σ^{70} Occupancy (Δ*hns*)


pstS
sbcD/phoB
phoA
ugpB
phoE/proB
phoH
phnC
rmf
agp
rpoH
cusR/cusC
yoaI/yeaL
feaR/feaB
mglB
wzaI/yegH
cspD/clpS
hemC/cyaA
csrA
ileS
rbsK
omrA
yhjJ
amn
glgB
frwA/frwC
yibT/ybiL
yhjB/yhjC
rhsE
(flhD)
(phoB)
(prp)
(lpxD)
(pmbA)
(proA)
(eptB)
(tktA)
(gloC)
(yahA)
(nudB)
(xylB)
(fpr)
(glnE)
(uxaC) uxaA
(yceA)
(yfiN) yfiB
(alaE)
(malF)
(yfcJ)
(ycjM)
(uacT)
(ydeN)
(metH)
(yhcA) yhcD
(ycdU)
(arsR)
(yddB)
(rhsJ)
(ybcK)
(evgS)
Escherichia
Enterobacter
Citrobacter
Yokenella
Salmonella
Raoultella
Cedecea
Klebsiella
Cronobacter
Hafnia
Dickeya
Yersinia
Rahnella
Pantoea
Erwinia
Brenneria
Pectobacterium
Edwardsiella
Providentia
Serratia
Proteus
Morganella
Leminorella
Photorhabdus
Xenorhabdus
Pseudomonas
Plesiomonas
Arsenophonus
Vibrio
Known Regulon
Intergenic Sites
Intragenic Sites
Weak
Strong
Predicted Site Strength