



Published as: *Cell*. 2012 July 6; 150(1): 222–232.

## Adaptive mutations that prevent cross-talk enable the expansion of paralogous signaling protein families

Emily J. Capra<sup>%</sup>, Barrett S. Perchuk<sup>%</sup>, Jeffrey M. Skerker<sup>#</sup>, and Michael T. Laub<sup>%,&,\*</sup>

<sup>%</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

<sup>&</sup>Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139

<sup>#</sup>Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720

### Summary

Orthologous proteins often harbor numerous substitutions, but whether these differences result from neutral or adaptive processes is usually unclear. To tackle this challenge, we examined the divergent evolution of a model bacterial signaling pathway comprising the kinase PhoR and its cognate substrate PhoB. We show that the specificity-determining residues of these proteins are typically under purifying selection, but have, in  $\alpha$ -proteobacteria, undergone a burst of diversification followed by extended stasis. By reversing mutations that accumulated in an  $\alpha$ -proteobacterial PhoR, we demonstrate that these substitutions were adaptive, enabling PhoR to avoid cross-talk with a paralogous pathway that arose specifically in  $\alpha$ -proteobacteria. Our findings demonstrate that duplication and the subsequent need to avoid cross-talk strongly influence signaling protein evolution. These results provide a concrete example of how system-wide insulation can be achieved post-duplication through a surprisingly limited number of mutations. Our work may help explain the apparent ease with which paralogous protein families expanded in all organisms.

### Introduction

The evolutionary forces and selective pressures that influence protein sequences remain poorly understood at a detailed, molecular level. A comparison of orthologs often reveals tens to hundreds of amino acid differences. How and why do functionally equivalent proteins diverge in different organisms? Many of the accumulated substitutions may be functionally neutral and result from processes such as genetic drift. However, some mutations may have been adaptive and provided a fitness advantage. Identifying these beneficial mutations and pinpointing the advantage that they provide are difficult problems. Comparative sequence analyses, such as measures of codon substitution patterns or dN/dS ratios (Yang and Bielawski, 2000), can help to identify residues that are potentially adaptive, but such approaches are frequently insufficient and difficult to validate. Additionally,

© 2012 Elsevier Inc. All rights reserved.

\*corresponding author: laub@mit.edu, 617-324-0418.

Supplemental information

Supplemental information consists of 4 figures and 4 tables.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

elucidating why certain mutations are beneficial requires a genetically manipulatable organism and an ability to probe the effects of individual mutations *in vivo*.

In many cases where protein evolution has been studied experimentally (reviewed in (Dean and Thornton, 2007)), the relevant proteins were examined *in vitro* or in heterologous hosts, and thus outside their native cellular context, possibly eliminating or obscuring important evolutionary constraints. For example, signal transduction proteins are often part of large paralogous families that expand through duplication and divergence. The duplication-divergence process thus runs an inherent risk of introducing cross-talk with existing pathways. A study of SH3 domains from *S. cerevisiae* and humans suggested that the avoidance of cross-talk may represent an important selective pressure in the evolution of paralogous protein families (Zarrinpar et al., 2003). However, a direct demonstration that cross-talk influences the evolution of signaling proteins and, more importantly, an understanding of how this occurs at the amino-acid level are lacking.

To tackle these challenges we examined the evolution of two-component signal transduction proteins in bacteria. These pathways, a primary means of signal transduction in prokaryotes, typically involve a sensor histidine kinase that, upon receipt of an input stimulus, autophosphorylates and then transfers its phosphoryl group to a cognate response regulator, which in turn modulates gene expression (Stock et al., 2000). Most histidine kinases are bifunctional and can, in the absence of an input signal, stimulate the dephosphorylation of their cognate response regulators, effectively acting as phosphatases (Huynh and Stewart, 2011).

Although most bacteria encode between 20 and 200 two-component signaling pathways (Alm et al., 2006), very little cross-talk occurs at the level of phosphotransfer *in vivo* (Grimshaw et al., 1998; Laub and Goulian, 2007; Siryaporn and Goulian, 2008; Skerker et al., 2005). Two-component pathways are highly specific, typically with one-to-one relationships between cognate kinases and regulators. When not stimulated to autophosphorylate, the phosphatase activity of a given histidine kinase can help to eliminate cross-talk and the errant phosphorylation of its cognate regulator (Siryaporn and Goulian, 2008). However, when stimulated as a kinase, molecular recognition is the dominant mechanism for preventing phosphotransfer cross-talk and thereby maintaining the fidelity of distinct signaling pathways. Systematic analyses of phosphotransfer have demonstrated that histidine kinases are endowed with an intrinsic ability to discriminate their *in vivo* cognate substrate from all other non-cognate substrates (Skerker et al., 2005). Analyses of amino acid coevolution in cognate signaling proteins identified the key specificity-determining residues in histidine kinases and response regulators (Bell et al., 2010; Capra et al., 2010; Casino et al., 2009; Skerker et al., 2008; Weigt et al., 2009). Rational mutagenesis of these residues can reprogram the partnering specificity of a histidine kinase or response regulator (Bell et al., 2010; Capra et al., 2010; Skerker et al., 2008).

For a given two-component pathway there is likely strong purifying selective pressure on its key specificity-determining residues to preserve the kinase-substrate interaction. Even single amino acid changes in specificity residues can drastically change the interaction capabilities and preferences of a histidine kinase or response regulator (Capra et al., 2010). Nevertheless, an inspection of orthologous kinases or response regulators often reveals divergent evolution and variability in the specificity residues of certain subsets of orthologs, raising the question of whether these changes resulted from neutral or adaptive processes. We favored the latter, hypothesizing that specificity residues must change in order to avoid cross-talk between pathways following gene duplication events. Two-component signaling pathways often expand through duplication (Alm et al., 2006), and following such events, bacteria presumably must accumulate mutations that insulate the new pathways from each other and

maintain their isolation from other, existing two-component pathways. Here, we provide direct experimental evidence that the avoidance of cross-talk is indeed a major selective force in the evolution of two-component signaling pathways. Through *in vitro* studies and fitness competition assays, we identify specific substitutions in a model two-component pathway, PhoR-PhoB, that represent an adaptation to the duplication of another two-component signaling system. Similar adaptations likely accompanied each of the duplication and divergence events underlying the massive expansion of two-component signaling protein families in bacteria. Accordingly, global analyses of specificity-determining residues in extant bacterial genomes reveal a pervasive trend toward orthogonality in these signaling proteins.

## Results

### Vertical inheritance of PhoR and PhoB

To examine the divergent evolution of two-component signaling pathways, we focused on the PhoR-PhoB signaling pathway (Wanner and Chang, 1987), which is found throughout the bacterial kingdom and helps a wide range of organisms respond to phosphate starvation. To systematically identify orthologs of *E. coli* *phoR* and *phoB*, we used a modified version of reciprocal best hits in BLAST analysis that allows for the identification of putative duplications. Most proteobacteria, except a small number of  $\delta$ -proteobacteria, were found to encode a single ortholog of each *phoR* and *phoB*, suggesting that these genes are rarely duplicated, particularly in the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -proteobacteria (Figure 1A). Additionally, gene trees for *phoR* and *phoB* closely matched a species tree (Figures 1B, S1A–B), indicating that this signaling system has likely been vertically inherited in these clades.

Given that *phoR* and *phoB* genes were rarely duplicated during the evolution of proteobacteria, it might be expected that the residues dictating phosphotransfer specificity would be relatively constant in order to preserve the interaction between PhoR and PhoB. We thus examined the six residues in PhoR and seven residues in PhoB previously identified as critical determinants of specificity in two-component signaling proteins (Capra et al., 2010). We extracted these residues, hereafter referred to simply as specificity residues, from 149 PhoR orthologs and 92 PhoB orthologs, and built sequence logos representing the relative frequency of amino acids at each specificity position (Figure 1B, Table S1). The difference in number of PhoR and PhoB orthologs results from the independent identification of kinase and regulator orthologs; most organisms encode both PhoR and PhoB.

The specificity residues of both PhoR and PhoB are generally well-conserved (Figures 1B, S1C), although several positions showed substantial variability. We then split the PhoR and PhoB sequences into groups corresponding to the three major proteobacterial subdivisions,  $\alpha$ ,  $\beta$ ,  $\gamma$ . Sequence logos built for each phylogenetic group revealed that differences between subdivisions can account for nearly all of the variability in the combined sequence logos (Figure 1B). For instance, in  $\gamma$ - and  $\beta$ -proteobacteria the first two positions are almost always threonine and valine, whereas in  $\alpha$ -proteobacteria these positions are usually alanine and serine or two alanines. Similar observations were made for the specificity residues of PhoB orthologs grouped according to phylogenetic subdivision. Importantly, each PhoR and PhoB sequence logo was built using species that are highly diverged. The strong conservation within each clade thus suggests that specificity residues are usually subject to strong purifying selection. Why, though, have specificity residues diverged between clades?

## Identification of adaptive mutations that prevent cross-talk *in vitro*

The clade-specific differences in PhoR and PhoB specificity residues may simply reflect degeneracy in the residues that enable PhoR and PhoB to interact. Alternatively, the differences may have produced functional changes such that a PhoR from one clade is less efficient at interacting with a PhoB from a different clade. To distinguish between these possibilities, we purified PhoR kinases from representative  $\gamma$ - and  $\alpha$ -proteobacteria, *E. coli* and *C. crescentus*, and examined their ability to phosphorylate a panel of 11 PhoB orthologs from  $\alpha$ ,  $\beta$ , and  $\gamma$ -proteobacteria (Figure 1C, S1D). For each PhoB from a  $\gamma$ -proteobacterium, phosphotransfer from the *E. coli* ( $\gamma$ ) PhoR was significantly faster than from *C. crescentus* ( $\alpha$ ) PhoR. Similarly, each PhoB from an  $\alpha$ -proteobacterium was preferentially phosphorylated by the  $\alpha$ -PhoR. For the two chosen  $\beta$ -PhoB orthologs, we observed more rapid phosphorylation by the  $\gamma$ -PhoR than the  $\alpha$ -PhoR, consistent with the specificity residues of the  $\beta$ -PhoR and  $\beta$ -PhoB orthologs being more similar to those found in  $\gamma$ -proteobacteria than those in  $\alpha$ -proteobacteria. We conclude that within each proteobacterial subdivision, the phosphotransfer specificity of PhoR and PhoB orthologs is relatively static. However, substitutions in the specificity residues of  $\alpha$ -PhoR and  $\alpha$ -PhoB orthologs have led to significant differences in phosphotransfer specificity between clades.

The changes in PhoR-PhoB specificity residues, and consequent alteration of interaction specificity, could have resulted from neutral drift. However, the strong conservation of specificity residues within each clade, which includes species that are widely divergent, suggests that such drift is extremely rare. Instead, the alternative PhoR-PhoB specificity residues in  $\alpha$ -proteobacteria may be adaptive and provide an important selective advantage. We hypothesized that the substitutions in  $\alpha$ -PhoR and  $\alpha$ -PhoB specificity residues prevent unwanted cross-talk with another pathway that is specific to the  $\alpha$ -proteobacteria, *i.e.* negative selection led to changes in  $\alpha$ -PhoR and  $\alpha$ -PhoB. This model predicts that PhoR orthologs from  $\gamma$ -proteobacteria may phosphorylate response regulators found exclusively in  $\alpha$ -proteobacteria, which the  $\alpha$ -PhoR orthologs have adapted to avoid phosphorylating.

To test this possibility, we performed comprehensive phosphotransfer profiling of *E. coli* ( $\gamma$ ) and *C. crescentus* ( $\alpha$ ) PhoR. Both PhoR constructs were autophosphorylated *in vitro* and then examined, in parallel, for phosphotransfer to the 44 response regulators encoded by *C. crescentus* (Figure 2). Both PhoR constructs phosphorylated the *C. crescentus* PhoB, consistent with their orthologous relationship; although as noted above, phosphotransfer from the  $\alpha$ -PhoR is more robust. Interestingly, the  $\gamma$ -PhoR showed significant phosphotransfer to NtrX, whereas the  $\alpha$ -PhoR construct did not. Notably, most  $\alpha$ -proteobacteria encode two paralogous Ntr systems, NtrB-NtrC and NtrX-NtrY, while the  $\gamma$ -proteobacteria typically encode only one, NtrB-NtrC (Figure 3A, Table S1). The two  $\alpha$ -Ntr systems, which likely arose through duplication and divergence, do not cross-talk with each other *in vitro* (Figure 3B) and, consistently, have different specificity residues (Figure 3C). Collectively, our observations suggest that the different PhoR specificity residues seen in  $\alpha$ -proteobacteria may have evolved to accommodate the presence of a second, lineage-specific pathway, NtrX-NtrY. Such a change in PhoR was presumably accompanied by changes in the PhoB specificity residues (see Figure 1B) to maintain phosphotransfer from PhoR.

Thus, we hypothesized that the alanine, serine, and phenylalanine found at specificity positions 1, 2, and 4 of  $\alpha$ -PhoR proteins represent adaptive mutations that prevent crosstalk to NtrX. To test this hypothesis, we created a series of *C. crescentus* PhoR mutants in which specificity residues were replaced with the corresponding residues from  $\gamma$ -proteobacterial PhoR. We made each single mutant, three double mutants, and the triple mutant. Each mutant kinase was then profiled against the complete set of *C. crescentus* response regulators to examine what effect, if any, these residues have on phosphotransfer specificity. Strikingly, each mutant led to a significant increase in NtrX phosphorylation (Figure 2).

We also examined detailed time courses of phosphotransfer from each mutant PhoR, as well as the wild-type kinases, to the *C. crescentus* regulators PhoB and NtrX. Each mutant kinase exhibited an increase in cross-talk with NtrX compared to the wild-type *C. crescentus* ( $\alpha$ ) PhoR, but retained the ability to phosphorylate *C. crescentus* PhoB at rates comparable to the wild-type PhoR (Figures 3D, S2A–C). Although some mutant PhoR kinases phosphorylated several substrates (see Figure 2), we focused on PhoB and NtrX as time-courses of phosphotransfer indicated these as the two preferred targets of mutant PhoR constructs (Figure S2D–E).

The most significant cross-talk to NtrX occurred for PhoR mutants with a valine substituted for serine at specificity position 2. Importantly, substantial cross-talk was not observed when this serine was substituted with other residues including leucine, aspartate, glutamate, and threonine. Only valine, corresponding to that found in  $\gamma$ -proteobacterial PhoR orthologs, produced significant cross-talk (Figure S2F–G).

Taken together, our *in vitro* studies support the notion that alanine, serine, and phenylalanine at specificity positions 1, 2, and 4 represent adaptive mutations that prevent cross-talk to NtrX in  $\alpha$ -proteobacteria.

### Avoidance of cross-talk is a significant selective pressure

To test whether these mutations also prevent cross-talk *in vivo*, we engineered the chromosomal copy of *phoR* in the  $\alpha$ -proteobacterium *C. crescentus* to produce a mutant PhoR in which specificity positions 1 and 2 are threonine and valine, respectively, as they are in  $\gamma$ -proteobacteria; hereafter this mutant strain is referred to as PhoR(TV). Based on our *in vitro* experiments, we expected that cross-talk from PhoR(TV) to NtrX would be induced during growth in phosphate-limited media (Figure 4A). During growth in such conditions, wild-type PhoR is stimulated to autophosphorylate and phosphotransfer to PhoB, which then activates genes involved in responding to phosphate limitation. Thus, any effects of increased cross-talk to NtrX by the PhoR(TV) kinase should be manifest specifically during growth in phosphate-limited media.

We grew cells to mid-logarithmic phase in phosphate-limited media and measured the rate of growth by monitoring the accumulation of optical density at 600 nm. In minimal media containing either 50  $\mu$ M phosphate or 5  $\mu$ M phosphate, the PhoR(TV) mutant grew significantly more slowly than wild type, with a doubling time ~30% longer than wild type in each case (Figures 4B, S3A–B). This growth defect was almost as severe as that observed for a  $\Delta$ *phoR* strain which cannot mount a proper transcriptional response to phosphate-limitation. To assess whether cross-talk from PhoR(TV) to NtrX contributed to the slow growth phenotype observed, we deleted *ntrX* in the PhoR(TV) strain. Indeed, the deletion of *ntrX* significantly reduced the growth deficiency of the PhoR(TV) mutant (Figures 4B, S3A–B) suggesting that cross-talk with NtrX contributes significantly to the slow growth phenotype of a PhoR(TV) strain. The suppression observed was not a non-specific acceleration of growth as the *ntrX* deletion alone had no effect on growth in phosphate-limited medium.

In phosphate-replete medium, the PhoR(TV) mutant strains grew at a rate nearly identical to the wild type (Figure 4B), indicating that, as expected, cross-talk to NtrX requires PhoR to be activated as a kinase. The *ntrX* deletion and PhoR(TV)/ $\Delta$ *ntrX* strains grew more slowly in phosphate-replete medium, as the NtrY-NtrX pathway is likely necessary for responding to a signal or metabolite produced in M2G medium.

To corroborate our growth rate measurements, we performed competitive fitness assays in which each mutant strain was mixed with the wild type at a ratio of 1:1 and grown in the

same flask for 104 hours, or approximately 40 wild-type generations. The mutant and wild-type strains were engineered to constitutively produce CFP or YFP, allowing for a rapid assessment of relative strain abundance using fluorescence microscopy. In phosphate-limited conditions, the PhoR(TV) strain showed a significant growth disadvantage, being almost completely eliminated from the population after 104 hours (Figures 4C, S3C). The fitness disadvantage of the PhoR(TV) mutant was comparable to that of  $\Delta phoR$  competed against wild type in the same phosphate-limited medium. Consistent with our growth measurements, deleting *ntrX* in the PhoR(TV) background improved competitive fitness (Figures 4C, S3C–D). In phosphate-replete conditions, the PhoR(TV) and  $\Delta phoR$  mutants retained a ratio with wild type close to 1:1, demonstrating that the selective disadvantage of introducing ancestral specificity residues into PhoR likely occurs only in conditions in which PhoR is a kinase. Collectively, these data further support a model in which the  $\alpha$ -specific substitutions in PhoR specificity residues (T→A and V→S at specificity positions 1 and 2) are selectively advantageous because they help prevent phosphotransfer cross-talk to NtrX, and perhaps other response regulators.

The growth and competitive fitness defects of PhoR(TV) in phosphate-limited media were comparable to that seen for  $\Delta phoR$ . This similarity suggested that the detrimental effect of cross-talk in the PhoR(TV) strain stems from an inability to phosphorylate PhoB and activate PhoB-dependent genes in phosphate-limited conditions. To test this hypothesis directly, we examined global gene expression patterns in the PhoR(TV) and  $\Delta phoR$  strains during growth in phosphate-limited conditions. These expression profiles exhibited strong similarity with a Pearson correlation coefficient of ~0.9 (Table S2), supporting a model in which phosphorylation cross-talk from PhoR(TV) to NtrX comes at the expense of phosphorylating PhoB. The inappropriate phosphorylation of NtrX could also contribute to the growth defect of the PhoR(TV) mutant. However, NtrX-dependent genes (see Experimental Procedures) were not significantly affected in the PhoR(TV) strain during growth in phosphate-limited conditions; NtrX-dependent genes behaved similarly in the PhoR(TV) and  $\Delta phoR$  strains in phosphate-limited conditions (Table S2). This may result from NtrY, the cognate kinase for NtrX, functioning as a phosphatase to prevent the accumulation of phosphorylated NtrX in phosphate-limited media. Consistent with this notion, *ntrX* and *ntrY* are not required for growth in phosphate-limited media, suggesting that in this condition NtrY is likely in a phosphatase state.

Importantly, and in contrast to NtrY, PhoR functions as a kinase, not a phosphatase, in phosphate-limited media. Thus, our results indicate that the  $\alpha$ -specific substitutions in PhoR specificity residues (T→A and V→S) impact fitness by affecting cross-talk at the level of phosphotransfer. Consistently, in phosphate-replete media, when PhoR is primarily active as a phosphatase, these substitutions had little to no effect on competitive fitness (Figure 4B–C, S3C–D). To further confirm that these substitutions do not significantly impact the phosphatase activity of PhoR, we examined global patterns of gene expression in the PhoR(TV) mutant grown in a phosphate-replete medium. Under these conditions, PhoR likely acts as a phosphatase to eliminate any errant phosphorylation of PhoB. Accordingly, the expression levels of known PhoB-dependent genes, such as *pstC*, *pstA*, and *pstB*, were modestly elevated in a  $\Delta phoR$  strain grown in phosphate-replete medium (Figure 4D). By contrast, these genes were not affected, or were slightly downregulated, in the PhoR(TV) strain grown in the same phosphate-replete conditions, indicating that PhoR(TV) retains phosphatase activity *in vivo*. Collectively, our data demonstrate that the growth and fitness defect of the PhoR(TV) mutant stems from inappropriate phosphotransfer to NtrX, and perhaps other non-cognate substrates.

### Different adaptive mutations prevent cross-talk in other proteobacterial clades

Our results suggest that  $\alpha$ -proteobacteria have accumulated substitutions in PhoR that prevent unwanted cross-talk with the non-cognate substrate NtrX. There could, however, be other ways to avoid cross-talk between these systems in other clades. Like the  $\alpha$ -proteobacteria, most  $\beta$ -proteobacteria encode NtrY-NtrX orthologs (Figure 3C). However, the  $\beta$ -PhoR orthologs have specificity residues at positions 1 and 2, similar to those found in  $\gamma$ -PhoR orthologs. This observation suggests that either the  $\beta$ -proteobacteria can tolerate cross-talk between PhoR and NtrX, or other mutations have emerged to prevent PhoR from phosphorylating NtrX. We favored the latter possibility as a comparison of sequence logos for the NtrX orthologs from  $\alpha$ - and  $\beta$ -proteobacteria revealed differences at two critical positions (Figure 3C). Whereas most  $\alpha$ -NtrX orthologs have aspartate, glycine, and lysine at specificity positions 2, 5, and 7, respectively, the  $\beta$ -NtrX orthologs typically have glycine, glutamate, and alanine at these same three respective positions. We speculated that the different specificity residues in a given  $\beta$ -NtrX may eliminate cross-talk from a  $\beta$ -PhoR; that is,  $\beta$ -proteobacteria may have evolved to avoid cross-talk by accumulating substitutions in NtrX rather than PhoR and PhoB.

To test this hypothesis, we asked whether introducing the  $\beta$ -NtrX specificity residues into an  $\alpha$ -NtrX would eliminate cross-talk from  $\alpha$ -PhoR(TV) which, as shown above, phosphotransfers to  $\alpha$ -NtrX *in vitro* and *in vivo*. Indeed, whereas *C. crescentus* NtrX was robustly phosphorylated by PhoR(TV), a mutant NtrX harboring the  $\beta$ -like substitutions D13G, G20E, F107I, and K108A was not detectably phosphorylated (Figure 4E). This mutant NtrX was not simply unfolded or unphosphorylatable as it was still robustly phosphorylated by  $\alpha$ -NtrY. Hence, the substitutions introduced specifically eliminated cross-talk from PhoR(TV), while still allowing for interaction with the cognate kinase NtrY. Taken together, these results suggest that in  $\beta$ -proteobacteria, substitutions in NtrX alleviated cross-talk with PhoR while in  $\alpha$ -proteobacteria substitutions in PhoR prevented cross-talk with NtrX. Although the substitutions are different, the net result in both cases was an insulation of the Ntr and Pho systems.

### Global optimization of signaling fidelity

Our results with the Pho and Ntr signaling pathways indicate that the avoidance of crosstalk following gene duplication is a major selective pressure that drives the accumulation of adaptive substitutions in the specificity-determining residues of two-component signaling proteins. More generally, this model predicts that the specificity residues of two-component signaling proteins in extant organisms should be sufficiently different from, or orthogonal to, one another to prevent cross-talk. To test this prediction, we extracted the six major specificity residues from each of the 22 canonical histidine kinases encoded in the *E. coli* K12 genome (Figure 5A). Pairwise comparisons indicated that kinases typically had no more than three identities with every other kinase at these six specificity sites, often with non-conservative differences at the remaining sites. One notable exception is NarX and NarQ, which contain two identities and four conservative differences. However, these kinases, which likely arose through gene duplication, each phosphorylate the response regulators NarL and NarP *in vitro* and likely *in vivo*, and hence represent a case of physiologically beneficial cross-regulation (Noriega et al., 2010). Aside from these two kinases, there is a general pattern of orthogonality between specificity residues in the system-wide set of *E. coli* histidine kinases. This orthogonality is further reflected by a lack of information in a sequence logo built from the specificity residues of the 22 *E. coli* histidine kinases (Figure 5B), particularly in comparison to the sequence logos built from orthologous histidine kinases (Figures 1B, 3C). A similar pattern of orthogonality was evident in the specificity residues of the 20 canonical histidine kinases in *C. crescentus*, as well as the specificity residues of the response regulators from both *E. coli* and *C. crescentus*.

(Figure S4). These observations, in combination with our detailed investigation of the Ntr and Pho proteins across phylogenies, suggest that the avoidance of cross-talk is a pervasive and significant selective pressure driving the system-wide insulation of two-component signaling pathways, and consequently, that in extant organisms, two-component systems are largely insulated from one another (Figure 5C).

## Discussion

Signaling protein families, in both prokaryotes and eukaryotes, frequently expand through gene duplication (Ohno, 1970; Pires-daSilva and Sommer, 2003). The retention of the duplicated genes often requires mutations that insulate them from one another, allowing each to transmit signals without inducing cross-talk. This divergence process may be additionally constrained by a need to avoid cross-talk with other, existing members of the same protein family. For two-component pathways, the duplication-divergence process can be conceptually framed by considering the sequence space defined by the specificity-determining residues of histidine kinases (Figures 5–6). For each kinase, these residues dictate the substrates it can phosphorylate, with different kinases recognizing largely non-overlapping, or orthogonal, sets of substrates. Gene duplication leads initially to a complete overlap and requires that one or both of the duplicates accumulate changes in its specificity residues, thus separating them in sequence space (Figure 6).

The mutational path taken by a given kinase may cause it to infringe on the sequence space occupied by another kinase, as was likely the case with the Ntr and Pho systems in  $\alpha$ -proteobacteria. Such overlap then necessitates additional mutations to achieve a system-wide optimization of specificity. Our results indicate that such optimization and the avoidance of cross-talk are important selective pressures influencing two-component systems and they can drive the divergent evolution of orthologous proteins.

How did the NtrY-NtrX pathway arise if cross-talk is detrimental? Although we cannot infer the order of events with complete certainty, a plausible scenario is that the NtrY-NtrX pathway arose during growth in phosphate-replete conditions where it provided a selective advantage, as suggested by the slow growth of a  $\Delta ntrX$  strain in these conditions. Subsequent growth in phosphate-limited conditions would then select for strains that have accumulated mutations eliminating cross-talk between PhoR and NtrX. We showed that such insulation can occur with only one or two point mutations, supporting the plausibility of this scenario in an ancestral  $\alpha$ -proteobacterium. Interestingly, the  $\beta$ -proteobacteria likely followed a different mutational path to avoiding cross-talk, accumulating substitutions in NtrX rather than PhoR. The difference between the two clades of proteobacteria may reflect the inherent stochasticity of mutations and selection. Alternatively, the growth conditions or genomic context of the ancestral organisms in which gene duplication occurred may have deterministically influenced selection.

In sum, we propose that the evolution of two-component signaling genes is characterized by long periods of stasis with specificity-determining residues subject to strong purifying selection to ensure robust phosphotransfer from kinase to regulator. Gene duplication, or lateral transfer events, can disrupt this stasis, requiring a global re-optimization of existing signaling proteins to accommodate the new pathway. The specificity residues are thus likely subject to bursts of diversifying selection; however, these residues would not necessarily exhibit commonly used signatures of diversifying selection such as large dN/dS values. Instead, our work emphasizes that a molecular-level understanding of protein evolution and the identification of adaptive mutations ultimately demands an integration of sequence analysis with focused biochemical and genetic characterizations.



Our approach and findings are relevant beyond two-component signaling as paralogous signaling protein families are found throughout biology. In fact, most organisms use a remarkably small number of types of signaling protein to carry out their diverse information-processing tasks. In all cases, duplication and divergence is a primary means by which new pathways are created and, consequently, issues of specificity and the fidelity of information transfer are critical. While eukaryotes sometimes rely on tissue-specific expression of paralogous genes or spatial mechanisms like scaffolds to enforce specificity, many common signaling proteins and domains, such as PDZ, SH3, SH2, and bZIP proteins (Hou et al., 2009; Liu et al., 2011; Newman and Keating, 2003; Stiffler et al., 2007; Tonikian et al., 2008; Zarrinpar et al., 2003), rely on molecular recognition and a relatively small set of specificity-determining residues. Hence, our observation that pathway insulation in bacteria can be achieved with a limited number of mutations may help to explain how organisms in all domains of life have exploited gene duplication to expand and diversify their signaling repertoires.

## Experimental Procedures

### Identification of orthologs and construction of gene trees

A modified version of reciprocal best blast hits was used to identify orthologous proteins. For *E. coli* PhoR, the DHP domain was used as a query in BLAST searches against fully sequenced bacterial genomes in GenBank (September 2009). The top ten hits from each genome were then subjected to reciprocal BLAST searches against the *E. coli* MG1655 genome. If only the top hit identified *E. coli* PhoR as the best match, it was called as a PhoR ortholog. If multiple hits identified *E. coli* PhoR as the best match, the top hit was called as an ortholog and additional hits were evaluated as follows. If an additional hit had an E-value within  $10^3$  of the top hit and was closer to the top hit than to the fifth hit (which generally had an E-value reflecting the overall paralogous relationship of histidine kinases), we also called it as an ortholog of *E. coli* PhoR and examined the next hit similarly. For genomes with more than one hit called as an ortholog, duplications were inferred and each hit deemed a member of the PhoR orthogroup. A similar procedure was followed to identify orthologs of PhoB, NtrB, NtrC, NtrX, and NtrY using as query sequences the *E. coli* PhoB receiver domain, the *C. crescentus* NtrB and NtrY DHP domains, or the *C. crescentus* NtrC and NtrX receiver domains.

Orthologous sequences were aligned using ClustalX (Chenna et al., 2003). Sequence logos for the specificity residues, extracted from the aligned sequences, were built using WebLogo (Crooks et al., 2004). To help correct for phylogenetic biases in genome sequencing efforts, sequences were filtered to ensure that no two sequences were more than 95% identical.

PhoR DHP domains and PhoB receiver domains were extracted using HMMER with models for a His\_KA domain or REC domain, respectively (Wistrand and Sonnhammer, 2005), and used to build gene trees through the PHYLIP package (Felsenstein, 1989) using the neighbor-joining algorithm provided. The tree was rooted using *B. subtilis* PhoR as the outgroup. Reported bootstrap values are out of 1000.

For genome-wide analyses of specificity residues, only canonical histidine kinases were included. Canonical kinases were defined as those containing the PFAM HisKA domain and no REC domain.

### Growth conditions and strain construction

*C. crescentus* cells were grown at 30°C in PYE, M2G (10 mM phosphate), M5G (50 μM phosphate), or M8G (same as M5G but with 5 μM phosphate), supplemented when necessary with oxytetracycline (1 μg/ml), kanamycin (25 μg/ml), 0.2% glucose or 0.3%

xylose. *E. coli* strains were grown at 37°C in LB supplemented with carbenicillin (100 µg/ml) or kanamycin (50 µg/ml). Transductions were performed using ΦCr30 (Ely, 1991).

Strains used are listed in Table S3. To construct ML1934, PhoR(TV), a region from nucleotide position –30 (relative to the PhoR start codon) to position 168 was amplified using CB15N genomic DNA as template and the primers PhoR\_int\_upstream\_for and PhoR\_int\_upstream\_rev, and a region from 147 to 1416 was amplified using pENTR-CC\_PhoR(TV) as template and the primers PhoR\_int\_for and PhoR\_int\_rev. Primer sequences are listed in Table S4. The two amplicons were then fused using SOE-PCR and ligated into pNPTS138, which had been cut with EcoRV and phosphatased using SAP, to create pNPTS-CC\_PhoR(TV). This plasmid was used for allelic replacement in CB15N following procedures described previously (Skerker et al., 2005). Integrants were tested for kanamycin sensitivity, sucrose resistance, and sequence-verified using primers listed in table S8. To create ML1935, PhoR(TV)/Δ*ntrX*, a tetracycline-marked *ntrX* deletion was transduced into ML1934; transductants were verified by PCR.

Strains used for competition assays (ML1936–ML1945) contained either the coding region for CFP or YFP driven by the *xyIX* promoter. *xyIX* was amplified from CB15N genomic DNA using primers *xyIX\_for* and *xyIX\_rev*, digested with KpnI and AgeI and ligated into pXCFPN-4 and pXYFPN-4 using the same restriction sites, producing pXCFPN-4:P<sub>*xyI*</sub>*cfp-xyIX* and pXYFPN-4:P<sub>*xyI*</sub>*yfp-xyIX*. The inserts were then amplified using primers *xyI\_xfp\_for* and *xyI\_xfp\_rev*, digested with HindIII and EcoRI, and ligated into pNPTS138 digested with the same enzymes. These vectors, pNPTS138:P<sub>*xyI*</sub>*cfp-xyIX* and pNPTS138:P<sub>*xyI*</sub>*yfp-xyIX*, were then integrated into the chromosomes of CB15N, Δ*phoR*, Δ*ntrX*, PhoR(TV), and PhoR(TV)/Δ*ntrX* through transformation and selection on kanamycin followed by counterselection on sucrose, leading to markerless integrations of CFP or YFP at the native *xyIX* locus of each strain.

Expression vectors were built by moving pENTR clones into destination vectors using the Gateway LR reaction (Invitrogen), and then transformed into BL21 *E. coli* for expression and purification. All site-directed mutagenesis was done on pENTR clones using primers listed in Table S4 and sequence-verified.

### Protein purification and phosphotransfer assays

Expression, protein purification, and phosphotransfer experiments were carried out as described previously (Skerker et al., 2005). Phosphotransfer profiles against all *C. crescentus* regulators comprise three gels, which were run in parallel and exposed to the same phosphorscreen. Gel images were then stitched together for presentation. Profiles used full-length response regulators except for CC1741 (*ntrC*), CC1743 (*ntrX*), and CC3743 (*cenR*) for which only receiver domains were used. For time courses of phosphotransfer in Figure 1C, each PhoB construct contained only the receiver domain.

### Growth and competitive fitness assays

Cultures were grown overnight in M2G. Cultures were then diluted to OD<sub>600</sub> ~ 0.025 and resuspended in either M2G or M5G. Samples were taken every hour and growth rates calculated 8–14 hours post-dilution to ensure phosphate-limitation of cells grown in M5G. For more severe phosphate limitation, growth curves were repeated in M8G, which contains 10-fold less phosphate. For these experiments, cultures were grown overnight in M5G and resuspended at an OD<sub>600</sub> ~ 0.07 in M8G.

For competitive fitness assays, cultures were grown overnight in M2GX, resuspended in either M2GX or M5GX to an OD<sub>600</sub> of 0.05, and mixed 1:1 with a competitor strain in 10 mL of media in a 150 mL flask. After 9 hours, a sample was taken and fixed using

paraformaldehyde. Cells were then diluted to  $OD_{600} \sim 0.01$  in 10 mL of M2GX or M5GX. After 15 hours, another sample was taken and cells diluted to  $OD_{600} \sim 0.05$ . After 9 hours, another sample was taken and cells diluted to  $OD_{600} \sim 0.01$ . This growth and dilution process was repeated until 104 total hours had elapsed. Cultures typically remained below  $OD_{600} \sim 0.85$  at all times. Cells from each sample collected were immobilized on 1.5% agarose pads made with PBS, and imaged using a Zeiss Axiovert 200 microscope with a 100x objective. Multiple fields of CFP, YFP, and phase images were taken for each sample. Roughly 500 cells were counted for each time point using a custom MATLAB script with counts checked manually using ImageJ. Competition experiments were done once with wild type expressing *cfp* and mutant expressing *yfp*, repeated with fluorescent proteins swapped, and results averaged.

### Microarray analysis

Cultures were grown to mid-log phase in M2G and either RNA was harvested or cells were washed, resuspended in M5G, and grown for 11 hours in phosphate-limited conditions before RNA was harvested. RNA was extracted, labeled, and hybridized to custom-designed 8×15K Agilent expression arrays as described previously (Gora et al., 2010). NtrX-dependent genes were defined as those genes exhibiting at least a 4-fold decrease in expression in the  $\Delta ntrX$  strain compared to wild-type *C. crescentus* in M2G. Complete array data are provided in Table S2 and deposited in GEO.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

We thank O. Ashenberg for help with bioinformatics, Y.E. Chen for help with strain construction, and A. Podgornaia, A. Keating, O. Ashenberg, and K. Foster for helpful comments on the manuscript. Sequence analyses were performed on a computer cluster supported by NSF grant 0821391. M.T.L. is an Early Career Scientist of the Howard Hughes Medical Institute. This work was supported by an NSF graduate fellowship to E.J.C and an NSF CAREER award to M.T.L.

### References

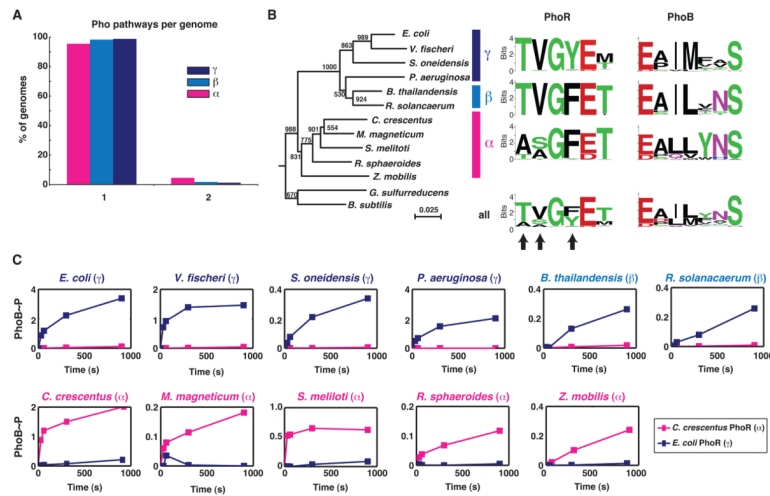
- Alm E, Huang K, Arkin A. The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. *PLoS Comput Biol.* 2006; 2:e143. [PubMed: 17083272]
- Bell CH, Porter SL, Strawson A, Stuart DI, Armitage JP. Using structural information to change the phosphotransfer specificity of a two-component chemotaxis signalling complex. *PLoS Biol.* 2010; 8:e1000306. [PubMed: 20161720]
- Capra EJ, Perchuk BS, Lubin EA, Ashenberg O, Skerker JM, Laub MT. Systematic dissection and trajectory-scanning mutagenesis of the molecular interface that ensures specificity of two-component signaling pathways. *PLoS Genet.* 2010; 6:e1001220. [PubMed: 21124821]
- Casino P, Rubio V, Marina A. Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. *Cell.* 2009; 139:325–336. [PubMed: 19800110]
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 2003; 31:3497–3500. [PubMed: 12824352]
- Crooks G, Hon G, Chandonia J, Brenner S. WebLogo: a sequence logo generator. *Genome Research.* 2004; 14:1188–1190. [PubMed: 15173120]
- Dean AM, Thornton JW. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet.* 2007; 8:675–688. [PubMed: 17703238]
- Ely B. Genetics of *Caulobacter crescentus*. *Methods Enzymol.* 1991; 204:372–384. [PubMed: 1658564]

- Felsenstein J. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*. 1989; 5:164–166.
- Gora KG, Tsokos CG, Chen YE, Srinivasan BS, Perchuk BS, Laub MT. A cell-type-specific protein-protein interaction modulates transcriptional activity of a master regulator in *Caulobacter crescentus*. *Mol Cell*. 2010; 39:455–467. [PubMed: 20598601]
- Grimshaw CE, Huang S, Hanstein CG, Strauch MA, Burbulys D, Wang L, Hoch JA, Whiteley JM. Synergistic kinetic interactions between components of the phosphorelay controlling sporulation in *Bacillus subtilis*. *Biochemistry*. 1998; 37:1365–1375. [PubMed: 9477965]
- Hou T, Xu Z, Zhang W, McLaughlin WA, Case DA, Xu Y, Wang W. Characterization of domain-peptide interaction interface: a generic structure-based model to decipher the binding specificity of SH3 domains. *Molecular & Cellular Proteomics*. 2009; 8:639–649. [PubMed: 19023120]
- Huynh TN, Stewart V. Negative control in two-component signal transduction by transmitter phosphatase activity. *Mol Microbiol*. 2011; 82:275–286. [PubMed: 21895797]
- Laub MT, Goulian M. Specificity in two-component signal transduction pathways. *Annu Rev Genet*. 2007; 41:121–145. [PubMed: 18076326]
- Liu BA, Shah E, Jablonowski K, Stergachis A, Engelmann B, Nash PD. The SH2 domain-containing proteins in 21 species establish the provenance and scope of phosphotyrosine signaling in eukaryotes. *Sci Signal*. 2011; 4:ra83. [PubMed: 22155787]
- Newman JR, Keating AE. Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science*. 2003; 300:2097–2101. [PubMed: 12805554]
- Noriega CE, Lin HY, Chen LL, Williams SB, Stewart V. Asymmetric cross-regulation between the nitrate-responsive NarX-NarL and NarQ-NarP two-component regulatory systems from *Escherichia coli* K-12. *Mol Microbiol*. 2010; 75:394–412. [PubMed: 19968795]
- Ohno, S., editor. *Evolution by Gene Duplication*. New York: Springer; 1970.
- Pires-daSilva A, Sommer RJ. The evolution of signalling pathways in animal development. *Nat Rev Genet*. 2003; 4:39–49. [PubMed: 12509752]
- Siryaporn A, Goulian M. Cross-talk suppression between the CpxA-CpxR and EnvZ-OmpR two-component systems in *E. coli*. *Mol Microbiol*. 2008; 70:494–506. [PubMed: 18761686]
- Skerker JM, Perchuk BS, Siryaporn A, Lubin EA, Ashenberg O, Goulian M, Laub MT. Rewiring the specificity of two-component signal transduction systems. *Cell*. 2008; 133:1043–1054. [PubMed: 18555780]
- Skerker JM, Prasol MS, Perchuk BS, Biondi EG, Laub MT. Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: a system-level analysis. *PLoS Biol*. 2005; 3:e334. [PubMed: 16176121]
- Stiffler MA, Chen JR, Grantcharova VP, Lei Y, Fuchs D, Allen JE, Zaslavskaja LA, MacBeath G. PDZ domain binding selectivity is optimized across the mouse proteome. *Science*. 2007; 317:364–369. [PubMed: 17641200]
- Stock A, Robinson V, Goudreau P. Two-component signal transduction. *Annual Review of Biochemistry*. 2000; 69:183–215.
- Tonikian R, Zhang Y, Sazinsky SL, Currell B, Yeh JH, Reva B, Held HA, Appleton BA, Evangelista M, Wu Y, et al. A specificity map for the PDZ domain family. *PLoS Biol*. 2008; 6:e239. [PubMed: 18828675]
- Wanner BL, Chang BD. The *phoBR* operon in *Escherichiacoli* K-12. *J Bacteriol*. 1987; 169:5569–5574. [PubMed: 2824439]
- Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci U S A*. 2009; 106:67–72. [PubMed: 19116270]
- Wistrand M, Sonnhammer E. Improved profile HMM performance by assessment of critical algorithmic features in SAM and HMMER. *BMC Bioinformatics*. 2005; 6:99. [PubMed: 15831105]
- Yamamoto K, Hirao K, Oshima T, Aiba H, Utsumi R, Ishihama A. Functional characterization in vitro of all two-component signal transduction systems from *Escherichia coli*. *J Biol Chem*. 2005; 280:1448–1456. [PubMed: 15522865]
- Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*. 2000; 15:496–503. [PubMed: 11114436]

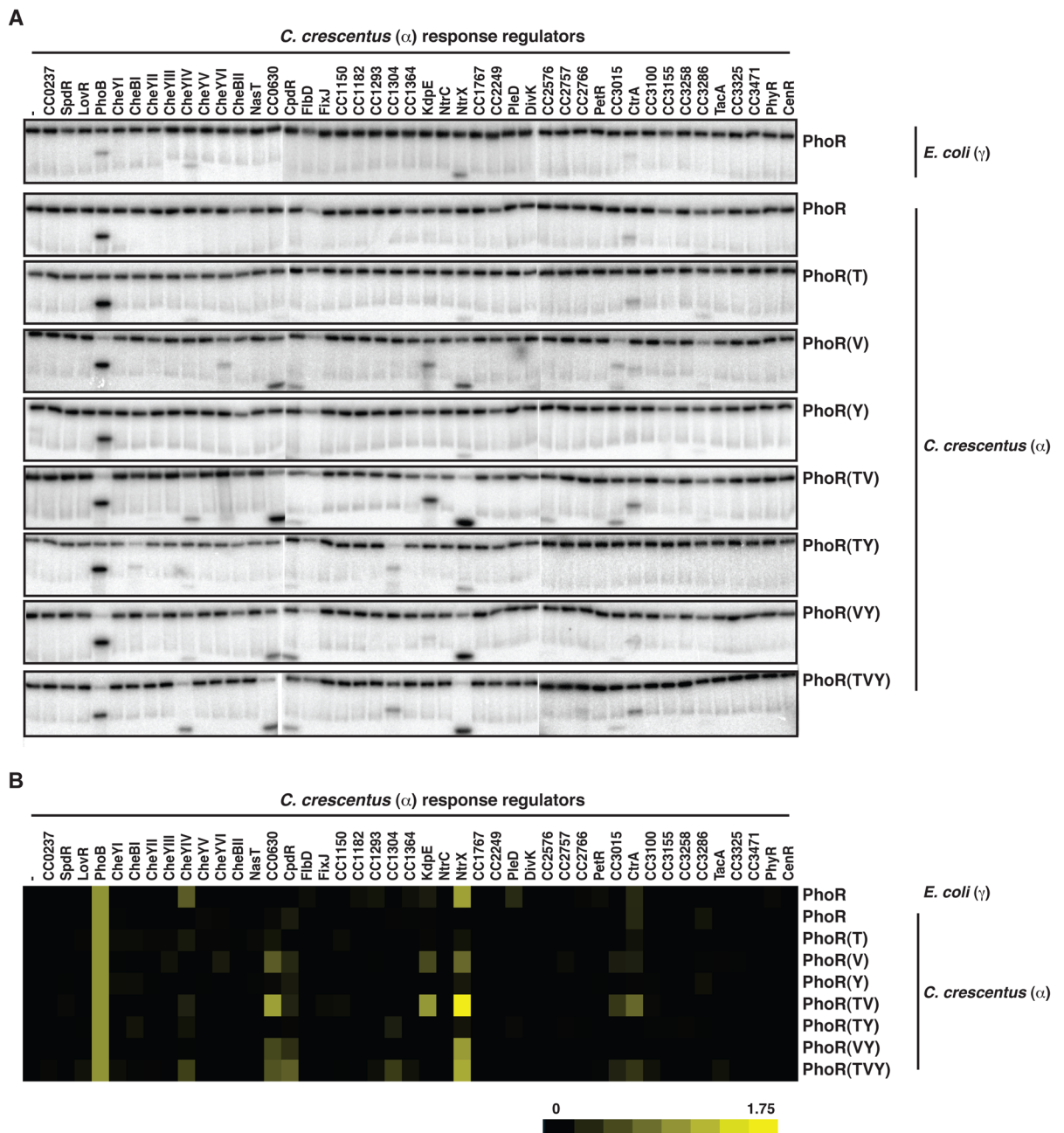
Zarrinpar A, Park SH, Lim WA. Optimization of specificity in a cellular protein interaction network by negative selection. *Nature*. 2003; 426:676–680. [PubMed: 14668868]

**Research Highlights**

the prevention of cross-talk constrains the evolution of signaling proteins  
unless duplicated, the specificity residues of signaling proteins rarely change  
avoiding cross-talk between pathways after duplication is a major selective pressure  
insulating duplicated signaling proteins requires changes in specificity residues

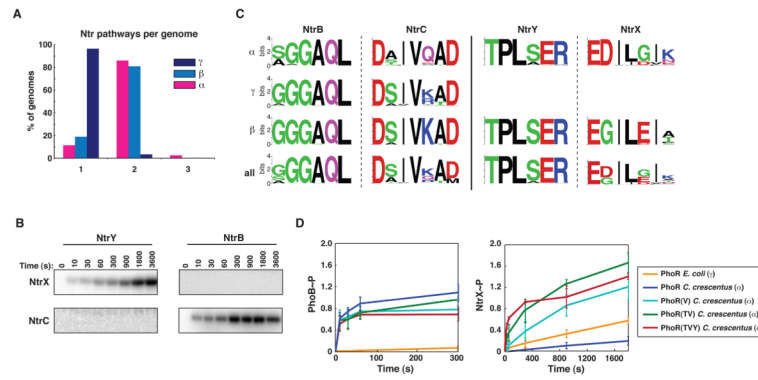


**Figure 1. Phosphotransfer specificity of PhoR is different in  $\alpha$ - and  $\gamma$ -proteobacteria**  
**(A)** Percentage of genomes harboring one or two Pho pathways. **(B)** Neighbor-joining tree for a subset of PhoR orthologs and sequence logos for specificity residues in PhoR and PhoB orthologs. Logos are shown for orthologs in each subdivision and as a combined set. Bootstrap values are out of 1000. A neighbor-joining tree for PhoB orthologs and a species tree for the species used are in Figure S1A–B. **(C)** Time-courses of phosphotransfer from *C. crescentus* PhoR ( $\alpha$ ) and *E. coli* PhoR ( $\gamma$ ) to each of 11 PhoB orthologs from representative  $\alpha$ -,  $\gamma$ -, and  $\beta$ -proteobacteria as noted above each graph. Band intensities for each PhoB were normalized in each experiment to the initial amount of autophosphorylated kinase. Values for PhoB phosphorylation can be greater than one as ATP was in excess in the reaction, allowing for re-autophosphorylation of PhoR and subsequent transfer to PhoB. For original gel images, see Figure S1D.



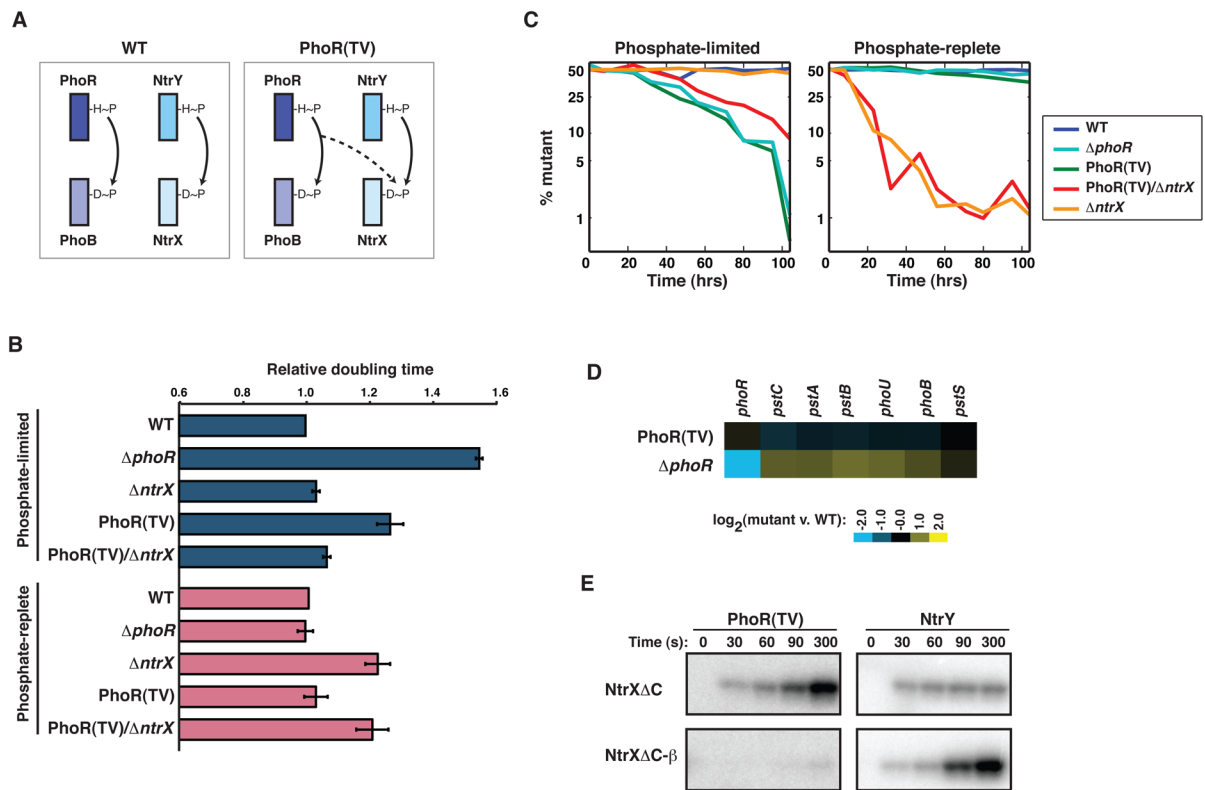
**Figure 2. Substituting  $\gamma$ -like specificity residues into  $\alpha$ -PhoR increases phosphorylation of NtrX** (A) Phosphotransfer profiling of *E. coli* PhoR, *C. crescentus* PhoR, and *C. crescentus* PhoR mutants containing  $\gamma$ -proteobacterial specificity residues. Each autophosphorylated kinase, indicated on the right, was incubated with each of 44 *C. crescentus* response regulators, indicated across the top, for 15 minutes. (B) Quantification of profiles in panel A. Band intensities for each response regulator were normalized to the level of autophosphorylated kinase and then plotted relative to PhoB. Also see Figure S2D–E.





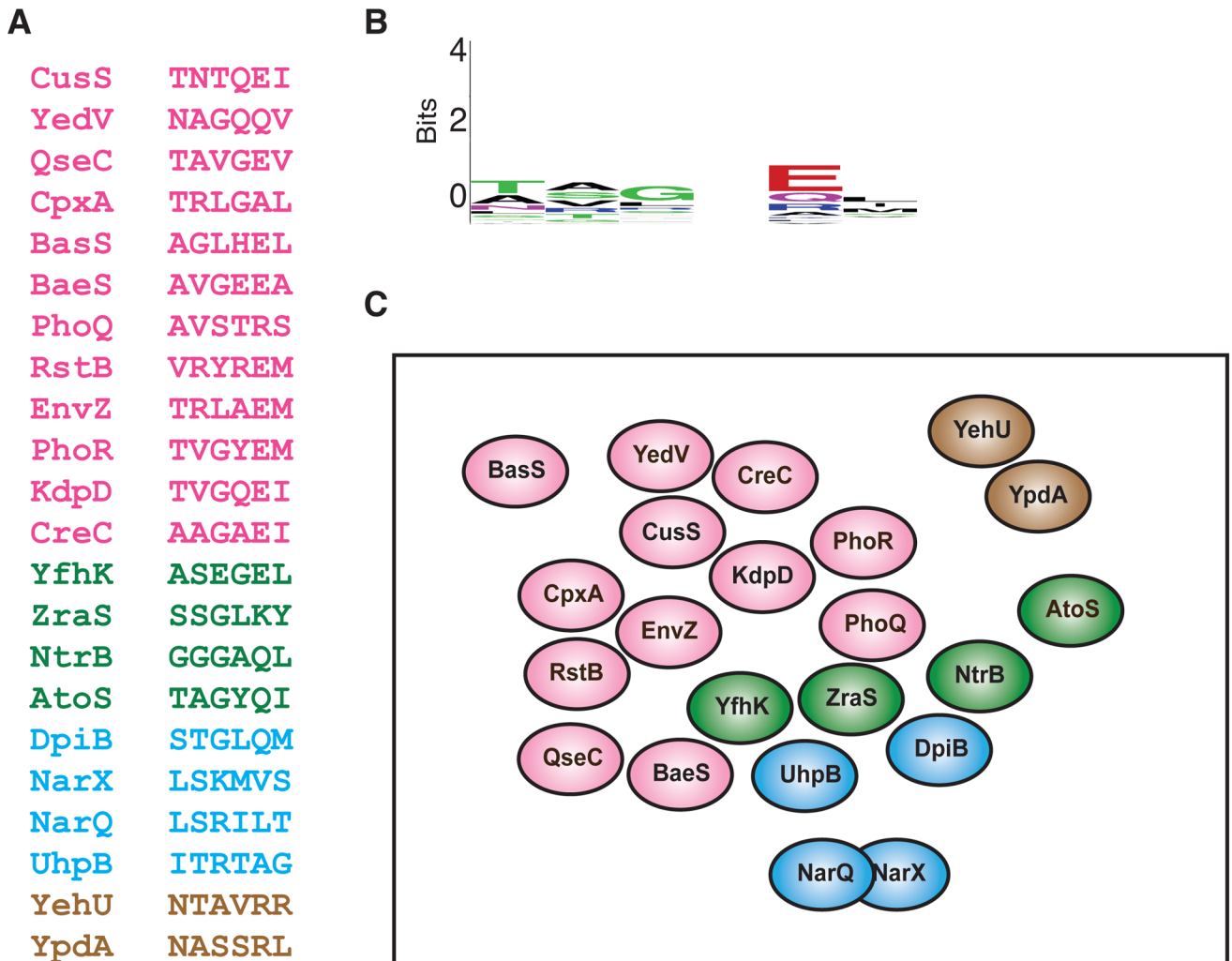
**Figure 3. The divergent evolution of NtrX after duplication led initially to cross-talk with PhoR in  $\alpha$ -proteobacteria**

(A) Percentage of genomes harboring one, two, or three Ntr pathways. (B) Time course of phosphotransfer from *C. crescentus* kinases NtrY and NtrB to *C. crescentus* response regulators NtrX and NtrC. The NtrY and NtrB constructs were autophosphorylated and examined for phosphotransfer at the time points indicated. Error bars represent standard deviations,  $n=3$ . (C) Sequence logos for specificity residues in NtrB-NtrC and NtrY-NtrX orthologs. Logos are shown for orthologs in each subdivision and as a combined set. (D) Time course of phosphotransfer from *E. coli* PhoR, *C. crescentus* PhoR, and *C. crescentus* PhoR mutants, listed in the legend, to *C. crescentus* PhoB and NtrX. Error bars represent standard deviations,  $n=3$ . Also see Figure S2A–C.



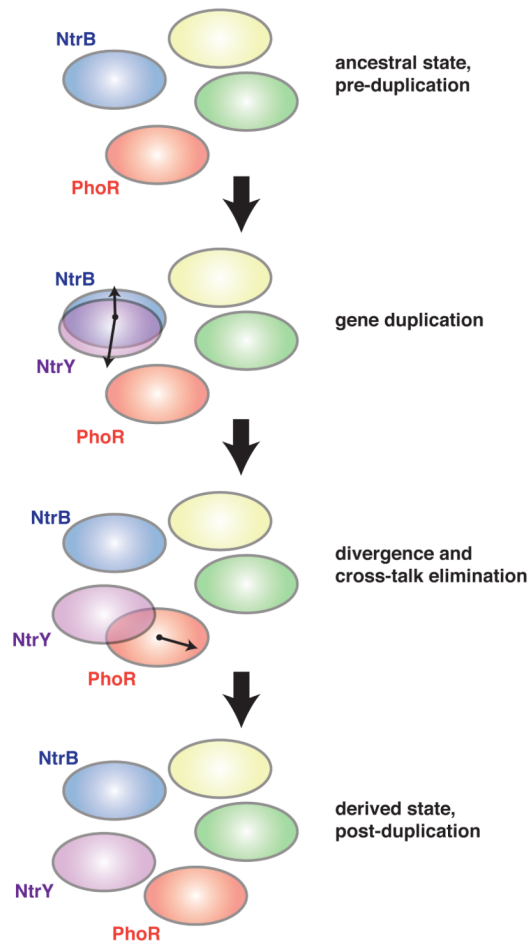
**Figure 4. Cross-talk between PhoR(TV) and NtrX leads to a growth defect and fitness disadvantage in phosphate-limited media**

(A) Schematic of strains examined. In the wild type, PhoR-PhoB and NtrY-NtrX are insulated, whereas the PhoR(TV) mutant leads to cross-talk between PhoR and NtrX. (B) Doubling times of *C. crescentus* strains  $\Delta phoR$ ,  $\Delta ntrX$ , PhoR(TV), and PhoR(TV)/ $\Delta ntrX$  relative to wild type in M5G (phosphate-limited) and M2G (phosphate-replete) media. Error bars indicate standard errors, n=3. For growth curves in M8G medium (5  $\mu$ M), see Figure S3A–B. (C) Wild type,  $\Delta phoR$ ,  $\Delta ntrX$ , PhoR(TV), and PhoR(TV)/ $\Delta ntrX$  were each competed against the wild type in M2GX and M5GX. The percentage of mutant cells in the population was measured periodically for 104 hours. Curves represent the average of two independent competitions with swapped fluorophores. Also see Figure S3C–D. (D) Expression data for known members of the *pho* regulon in PhoR(TV) and  $\Delta phoR$  in M2G (phosphate-replete) media. Data are expressed as  $\log_2$  values of the ratio between a given mutant and wild-type *C. crescentus*, and are color-coded according to the legend. (E) Time courses of phosphotransfer from kinases PhoR(TV) and NtrY to the regulators NtrX $\Delta$ C and an NtrX $\Delta$ C harboring  $\beta$ -like specificity substitutions. Only the response regulator band is shown.



**Figure 5. Extant two-component signaling pathways are insulated from each other at the level of phosphotransfer**

(A) The six primary specificity residues are shown for each of the 22 canonical *E. coli* histidine kinases. Hybrid histidine kinases and the non-canonical kinases DcuS and CheA were omitted. The histidine kinases are separated into groups by color based on the family of their cognate response regulator: pink, OmpR/winged helix-turn helix; green, NtrC/AAA + and Fis domains; blue, NarL/GerE helix-turn-helix; brown, LytR. For specificity residues from *E. coli* response regulators and *C. crescentus* histidine kinases and response regulators, see Figure S4. (B) Sequence logo for the specificity residues in panel A. (C) A qualitative two-dimensional representation of the distribution of *E. coli* histidine kinases in the sequence space defined by the six primary specificity-determining residues. Each oval represents the set of response regulators recognized by a histidine kinase given its specificity residues. Spheres are colored using the same scheme as in panel A. With the exception of NarQ and NarX (see text), the spheres are non-overlapping, indicating a lack of cross-talk *in vivo* and *in vitro*. Kinases were placed relative to one another based roughly on their ability to phosphorylate the cognate regulators of other histidine kinases after extended incubation times *in vitro* (Skerker et al., 2005; Yamamoto et al., 2005). For example, CpxA shows a strong preference for phosphotransfer to its cognate regulator CpxR, but will phosphorylate the cognate regulators of EnvZ and RstB after extended periods of time.



**Figure 6. Adaptive divergence of duplicated signaling pathways involves the elimination of cross-talk**

Ovals represent the set of response regulators recognized by a histidine kinase, as determined by its specificity residues (see also Figure 5). The NtrB-NtrC pathway is shown duplicating to produce the paralogous system NtrY-NtrX. As these pathways diverged, the specificity of NtrY overlapped that of PhoR, necessitating a change in PhoR specificity to yield the derived state with insulated pathways.