Microbiology (2010), **156**, 14–22 DOI 10.1099/mic.0.033266-0

Review

New insights into the regulatory networks of paralogous genes in bacteria

Mario A. Martínez-Núñez,[1] Ernesto Pérez-Rueda,[2] Rosa María Gutiérrez-Ríos[1] and Enrique Merino[1]

Correspondence
Enrique Merino
merino@ibt.unam.mx

[1]Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

[2]Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

Extensive genomic studies on gene duplication in model organisms such as *Escherichia coli* and *Saccharomyces cerevisiae* have recently been undertaken. In these models, it is commonly considered that a duplication event may include a transcription factor (TF), a target gene, or both. Following a gene duplication episode, varying scenarios have been postulated to describe the evolution of the regulatory network. However, in most of these, the TFs have emerged as the most important and in some cases the only factor shaping the regulatory network as the organism responds to a natural selection process, in order to fulfil its metabolic needs. Recent findings concerning the regulatory role played by elements other than TFs have indicated the need to reassess these early models. Thus, we performed an exhaustive review of paralogous gene regulation in *E. coli* and *Bacillus subtilis* based on published information, available in the NCBI PubMed database and in well-established regulatory databases. Our survey reinforces the notion that despite TFs being the most prominent components shaping the regulatory networks, other elements are also important. These include small RNAs, riboswitches, RNA-binding proteins, sigma factors, protein–protein interactions and DNA supercoiling, which modulate the expression of genes involved in particular metabolic processes or induce a more complex response in terms of the regulatory networks of paralogous genes in an integrated interplay with TFs.

Introduction

Gene duplication is one of the main sources of functional divergence in organisms (Babu *et al.*, 2004; Conant & Wolfe, 2008; Gelfand, 2006; Lynch & Conery, 2000; Teichmann & Babu, 2004). In order to fulfil an organism's metabolic needs, selection processes modify the regulation of the paralogous gene copies, taking advantage of a repertoire of new functions. Duplication events may include genes coding for transcription factors (TFs), permitting a more versatile adaptation of the functional diversity gained from the duplication of structural genes. Different aspects of the evolution of the regulatory networks of paralogous genes have been examined, including the co-evolution of the upstream regulatory regions and their corresponding TFs; the likely consequences of gain, loss, and replacement of TFs in the regulatory networks of paralogous genes (Gelfand, 2006; Teichmann & Babu, 2004); and also the topological and dynamic properties of the regulatory networks (Babu *et al.*, 2006; Balaji *et al.*, 2007; Luscombe *et al.*, 2004). This review will consider the fascinating repertoire of additional mechanisms other than TFs, which help regulate the expression of paralogous genes in *Escherichia coli* and *Bacillus subtilis*. This analysis is derived from an extensive compilation of the regulatory information available from the NCBI PubMed database and deposited in the *E. coli* RegulonDB Release 6.3 (Gama-Castro *et al.*, 2008), *B. subtilis* DBTBS (Sierro *et al.*, 2008) and Rfam version 9.1 (Gardner *et al.*, 2009) databases.

Four supplementary tables are available with the online version of this paper.

TFs are essential elements in the regulatory networks of paralogous genes

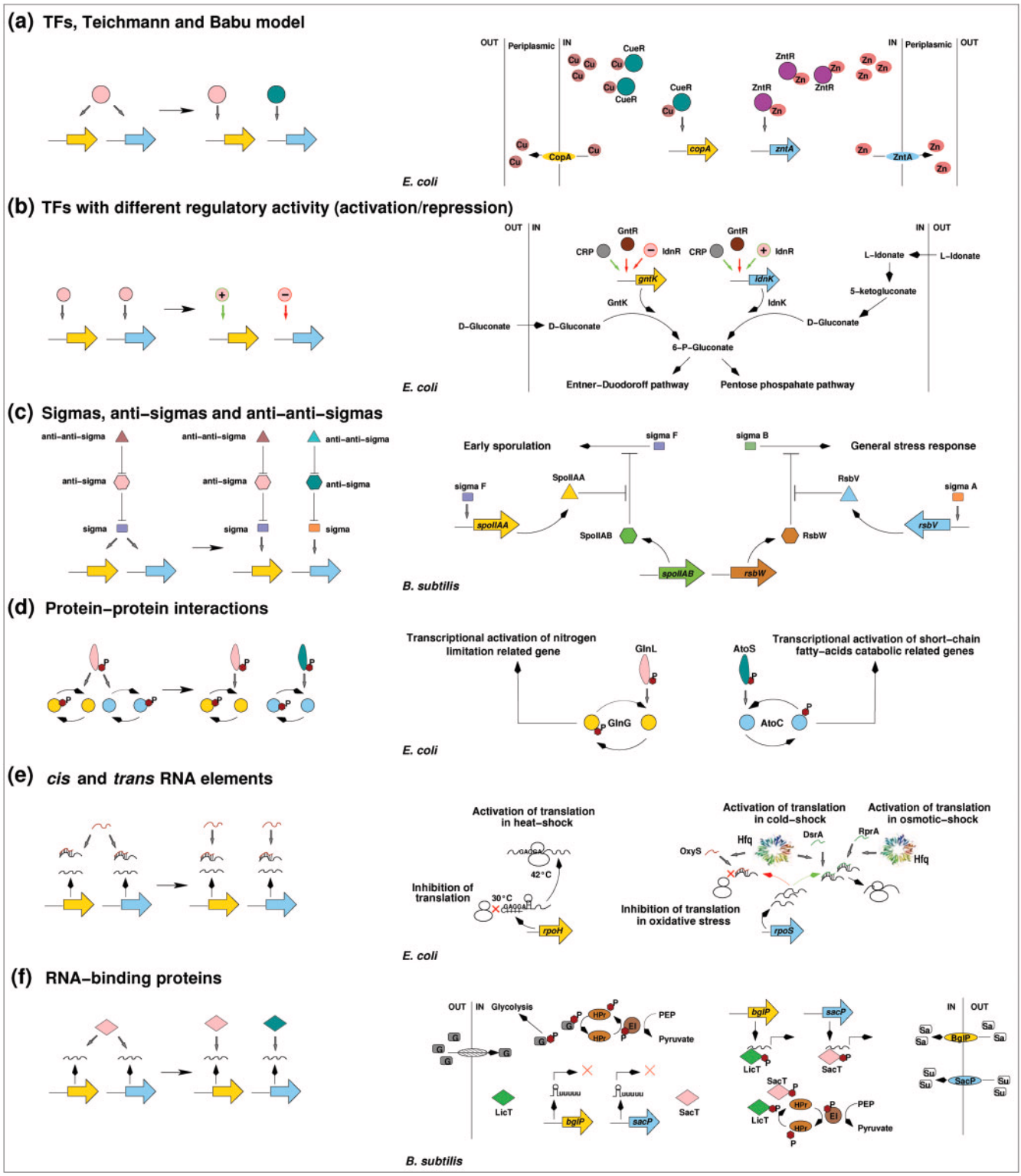
TFs are the most prominent and usually the only regulatory element taken into account in any of the aforementioned studies because of the important role they play in the regulatory networks of paralogous genes. Teichmann & Babu (2004) formulated a widely accepted evolutionary model portraying the regulation of duplicated genes. In this model, the loss and gain of regulatory interactions may occur following the duplication of either a TF or a target gene, or following the duplication of both a TF and a target gene (Fig. 1a). The Teichmann and Babu model does not

question whether the regulatory effect (either activation or repression) of the TFs on the paralogous genes is conserved or changed in any of these instances; it only considers the regulatory interactions. An example of a regulatory conserving effect is found in the *E. coli* paralogous genes *zntA* and *copA* (Group 393, Supplementary Tables S1 and S3), coding for zinc and copper transporters, which are regulated by the paralogous TFs ZntR and CueR, respectively (Yamamoto & Ishihama, 2005a, b) (Fig. 1a). We also identified examples of regulatory systems where the TF is maintained but a different effect is gained (either activation or repression) in one of the paralogous genes. This evolutionary scenario can be observed in the regulation of the *E. coli gntK* and *idnK* gluconate kinase genes, involved in the 6-phosphogluconate synthesis which takes place in the Entner–Doudoroff and pentose phosphate pathways, respectively. Although these genes are regulated by the same TFs, CRP, GntR and IdnR, the last TF represses the transcription of *gntK*, whereas it activates the transcription of *idnK* (Bausch *et al.*, 2004; Izu *et al.*, 1997) (see Fig. 1b and Group 449 of Supplementary Tables S1 and S3).

In order to quantify the divergence in the regulation of paralogous genes in *E. coli* and *B. subtilis*, BLAST comparisons (Altschul *et al.*, 1990) of their corresponding protein sequence sets were carried out. Based on these comparisons, we identified 477 groups of paralogous genes in *E. coli*. From this set, 291 groups of paralogous genes were found to be regulated by at least one TF, and for 43 of these (14.8 %) the TFs were shown to have a dual activation/repression effect on their corresponding paralogous target genes (Supplementary Table S3). In *B. subtilis* this tendency was considerably less. Out of the 483 groups of paralogous genes, we found that 225 groups were regulated by at least one TF, and eight of these (3.6 %) presented at least one TF with a dual activation/repression regulatory activity (Supplementary Table S4). In *E. coli*, this dual regulatory activity is more frequently found in the groups of paralogous genes regulated by global TFs: ArcA, CRP, FIS, FNR, Hns, IHF and Lrp (Martinez-Antonio & Collado-Vides, 2003) (see Supplementary Table S3). In contrast, only some of the *B. subtilis* global TFs such as AbrB, CcpA, Spo0A and TnrA (Moreno-Campuzano *et al.*, 2006) exhibit a dual activation/repression activity, whereas others either act always as repressors (CodY, Fur, LexA and YrzC), or always as activators (ComK) (see Supplementary Table S4). Finally, a smaller proportion of local TFs show dual activation/repression regulatory activity on paralogous genes. This is the case for the *E. coli* NtrC transcription regulator, which represses the transcription of the glutamine transporter subunit *glnH* gene and activates the transcription of the histidine/lysine/arginine/ornithine transporter *hisJ* and the lysine/arginine/ornithine transporter *argT* genes (Zimmer *et al.*, 2000) (Group 27 of Supplementary Table S1).

In order to test the statistical significance of finding a regulatory element (TFs, sigma factors, protein–protein interactions, DNA supercoiling, RNAs and RNA-binding proteins) that regulates two or more paralogous genes in a regulatory network, we carried out a computer simulation procedure, similar to the one performed by Teichmann & Babu (2004), where the regulatory interactions were randomly shuffled. Having performed the randomization process 10 000 times, we compared the mean values of these simulations against the number of interactions in the real regulatory networks of *E. coli* and *B. subtilis* (531 and 189 interactions, respectively). These real interactions were found to be significantly different from the average interactions of their corresponding randomized *E. coli* and *B. subtilis* datasets (19.5 and 19 standard deviations, SDs, respectively). The statistical significance of two or more paralogous genes being regulated by a similar TF, but having a dissimilar regulation (activation vs repression), was also evaluated. In this case, the distance of the real values from the mean of randomly generated networks in *E. coli* and *B. subtilis* was 3.7 and 2 SDs, respectively. Interestingly, when this analysis was performed excluding the global regulators of *E. coli* (ArcA, CRP, FIS, FNR, Hns, IHF and Lrp) (Martinez-Antonio & Collado-Vides, 2003) and *B. subtilis* (AbrB, ComK, TnrA, CodY, Spo0A and CcpA) (Moreno-Campuzano *et al.*, 2006), both distances increased significantly, to 6.5 and 3.4 SDs, respectively. These results highlight the evolutionary plasticity of the regulatory networks: not only as a result of the duplication of TFs' interactions on the regulatory network as previously proposed by Teichmann & Babu (2004), but also as a result of the divergent effect of the TF interactions, activating or repressing the transcription of paralogous genes.

Divergence of paralogous genes may include the modification or acquisition of new regulatory mechanisms, changes in gene dosage (Gu *et al.*, 2002), subdivision of ancestral functions and the evolution of new functions (Conant & Wolfe, 2008). Our results suggest that modulation is one of the first steps towards evolutionary innovation at a biochemical level, perhaps as a step towards the modification of the entire pathway. In this regard, diverse authors have proposed that while paralogues may retain similar functionality, gene expression rapidly diverges immediately after duplication events (Gu *et al.*, 2005, 2002; Li *et al.*, 2005), implying that alterations in gene expression precede potential changes in function. Furthermore, it has been postulated that paralogous genes which diverge in function contribute to evolutionary innovation at a biochemical level (Evangelisti & Wagner, 2004; Gu *et al.*, 2005, 2002; Zhang *et al.*, 1998). In order to corroborate these observations, we analysed the metabolic divergence of paralogous genes. By identifying the differences in the assigned biochemical reaction or metabolic pathway, we were able to show that 198 (42 %) of the paralogous groups in *E. coli* have a known metabolic function, according to the KEGG database (Gu *et al.*, 2005, 2002; Kanehisa *et al.*, 2008). From this set, we identified 53 groups (27 %) that possess members which share the same metabolic pathway



and perform the same enzymic reactions; 64 groups (32 %) which are formed exclusively from proteins that belong to different pathways; and 26 groups (13 %) which are entirely constituted of proteins that are part of the same pathway, although they take different metabolic steps. Intermediate cases of groups partially formed by any of the above categories were found on 55 occasions (28 %). Slightly different tendencies were observed in the 485 groups of paralogous genes in *B. subtilis*. Finally, duplication events may not be restricted to individual

Fig. 1. Regulatory elements involved in the regulation of the paralogous gene copies. (a) Simplification of the model of Teichmann & Babu (2004), where a new TF is gained in order to regulate one of the paralogous gene copies. TFs are represented by circles. (b) Extension of the Teichmann and Babu model where the differential regulation of the paralogous genes depends on the different regulatory activity (either activation or repression) of a TF on its gene targets. The activation and repression activities of the TFs are represented by circles with green and red perimeters, respectively. (c) Regulatory networks of paralogous genes dependent on sigma, anti-sigma and anti-anti-sigma factors. These factors are represented by rectangles, hexagons and triangles, respectively. The inhibitory activity of these elements is represented by T-like lines. (d) Regulatory networks of paralogous genes dependent on protein–protein interactions. In bacteria, two-component signal-transduction systems are one of the most common types of regulatory networks. The histidine kinases of the system are represented by ovals, and the response regulators by circles. (e) Regulatory networks of paralogous genes dependent on *cis*- and *trans*-RNA elements. *cis*-RNA elements are represented by stem-and-loop structures, whereas different small RNAs are represented by short undulating lines. (f) Regulatory networks of paralogous genes dependent on RNA-binding proteins (represented by rhomboids).

genes, but may also involve large parts of metabolic pathways, as is the case with the leucine, arginine and lysine biosynthesis pathways, where several enzymes are involved in related chemical reactions, with analogous but different substrates (Fondi *et al.*, 2007).

Alternative sigma, anti-sigma and anti-anti-sigma factors may also be part of the regulatory network of paralogous genes

Transcription regulation by alternative sigma factors, anti-sigma factors and anti-anti-sigma factors is a common strategy used by various organisms in order to accomplish different cellular processes, including (i) the regulation of specific biogenesis processes, such as flagellar biosynthesis in *E. coli*, which is regulated by σ^{28} (Chilcott & Hughes, 2000); (ii) the response to a metabolic requirement, as occurs when *E. coli* grows in nitrogen-deprived conditions and gene transcription depends on σ^{54} (Reitzer & Schneider, 2001); (iii) the response to different kinds of stress factors, as in heat shock stress in *E. coli*, which is regulated by σ^{32}, or in the case of general stress in *B. subtilis*, which is regulated by σ^{B} (Hecker *et al.*, 2007); (iv) synchronized transcription at a specific growth stage, as occurs in the stationary phase of *E. coli*, where transcription is regulated by σ^{38} (Hengge-Aronis, 2002); and (v) coordinated cell differentiation processes, such as *B. subtilis* forespore formation, whose regulation depends on genes transcribed by σ^{E}, σ^{F}, σ^{G}, σ^{H} and σ^{K} (Kroos & Yu, 2000). Although in general each of the sigma factors recognizes clearly distinguishable promoter sequences and thus transcribes different groups of genes, in *E. coli* the stationary-phase σ^{38} and its paralogous counterpart, the housekeeping σ^{70}, may share overlapping sets of target genes (Typas *et al.*, 2007). The presence of *cis*-acting elements located in the promoter region, as well as the activity of *trans*-acting proteins, establishes whether a promoter is recognized by RNA polymerase containing σ^{38} or σ^{70}. This transcription regulation based on σ^{38} and σ^{70} constitutes a clear example of an extension of the Teichmann and Babu model, where the regulatory network grows by the duplication of regulatory elements other than TFs, which modulate the expression of common gene targets.

According to our compiled data referring to the regulatory networks of paralogous genes, there are 100 out of 477 (21 %) groups in *E. coli* and 144 out of 483 (30 %) groups in *B. subtilis* that are regulated by alternative sigma, anti-sigma or anti-anti-sigma factors (Supplementary Tables S1 and S2). In order to evaluate the statistical significance of finding a sigma factor which regulates two or more paralogous genes, we repeated the simulation procedure performed in the TF analysis, using the sigma factor interaction as input data. Our statistical analysis shows that these sigma factor interactions on paralogous genes in *E. coli* and *B. subtilis* were highly significant, as they registered 18 and 6.8 SDs above the mean shown by those obtained in the corresponding randomized regulatory networks. The regulatory model of paralogous genes based on sigma factors is illustrated by the stress response and the early sporulation stage of *B. subtilis,* which both depend on the paralogous sigma factors σ^{B} and σ^{F}. The regulatory network of these systems also involves a set of paralogous anti-sigma and anti-anti-sigma factors and represents an excellent example of an extension of the Teichmann and Babu model, where the regulatory elements are not TFs. In a first instance, the transcriptional role of σ^{B} and σ^{F} on their target genes is inhibited by the binding of the anti-sigma factors RsbW and SpoIIAB (Group 232, Supplementary Tables S2 and S4), respectively. In a second instance, these paralogous anti-sigma factors are negatively regulated by the anti-anti-sigma factors RsbV and SpoIIAA (Group 313, Supplementary Tables S2 and S4). Finally, the transcription of the genes that encode these paralogous anti-anti-sigma factors depends on the paralogous sigma factors σ^{A} and σ^{F} (Hecker *et al.*, 2007) (Fig. 1c).

The impact of evolutionary forces on the topological structures known as motifs has been analysed for different kinds of networks (Shen-Orr *et al.*, 2002); for example biological regulatory networks (Kashtan & Alon, 2005; Shen-Orr *et al.*, 2002) and paralogous gene networks (Teichmann & Babu, 2004). It has been reported that duplication of an entirely feed-forward motif (a topological

structure where a TF regulates a second TF, and both TFs simultaneously regulate a target gene) has not been observed in the regulatory networks of model organisms, although single genes generated by duplication could be part of a new feed-forward or other kind of motif. For example, in *B. subtilis* the paralogues σ^F and σ^E are part of different feed-forward motifs. In the first case, σ^F forms a feed-forward motif with the anti-anti-sigma factor SpoIIAA and the anti-sigma factor SpoIIAB, as σ^F regulates the expression of SpoIIAA and SpoIIAB, whereas SpoIIAA modulates the expression of SpoIIAB. In a similar manner, the second feed-forward loop is formed by σ^E, PhoP and PhoR, which are involved in phosphate uptake, post-exponential growth phase and other stress responses (Pragai *et al.*, 2004). Duplication events, or mutations in the duplicated regulatory genes or in the regulatory target sites, can generate new feed-forward motifs useful for the rewiring of the regulatory networks of the paralogous genes, which favours the adaptation process of the organism as it responds to niche adaptation (Gelfand, 2006).

Protein–protein interactions modulate the activity of paralogous proteins

The regulatory networks evolved by gene duplication are not exclusively regulated at the level of transcription initiation. The activity of paralogous proteins can also be significantly post-translationally modified by protein–protein interactions. This kind of regulation is mainly represented by the two-component signal-transduction systems. Most of these systems consist of a membrane-bound histidine kinase that senses a specific environmental signal and catalyses the transfer of a phosphate group to its corresponding response regulator, inducing its activation. An example of this regulation is found in *E. coli* by the paralogous TFs GlnG and AtoC, involved in the transcriptional activation of genes in response to nitrogen limitation and the catabolism of short-chain fatty acids, respectively. The regulatory activity of these TFs depends on their corresponding paralogous histidine kinases GlnL and AtoS (Lioliou *et al.*, 2005; Ninfa & Magasanik, 1986) (see Fig. 1d and Group 8 of Supplementary Tables S1 and S3). Another well-characterized example of the regulation of paralogous proteins by protein–protein interaction is observed in the chemotactic response of *E. coli*. In this case CheR and CheB cause the post-translational regulation of the methyl-accepting chemotaxis proteins Trg, Tap, Tar, Aer and Tsr, paralogous proteins that undergo changes in their methylation state in response to chemical stimuli, thus establishing the swimming pattern (Ferris & Minamino, 2006) (See Group 41 of Supplementary Tables S1 and S3). From our set of paralogous genes, we found nine groups regulated by protein–protein interactions in *E. coli* and nine groups in *B. subtilis* (Supplementary Tables S1 and S2). The numbers of post-translational interactions that involve one regulator modulating the activity of two or more paralogous genes in *E. coli* and *B. subtilis* were 16 and 12, respectively. Despite the fact that these interactions represent only a small fraction of their corresponding regulatory networks, we found these values to be statistically significant, as they are 23.5 and 15.5 SDs above the means of the corresponding *E. coli* and *B. subtilis* randomly generated networks. It is worth mentioning that these protein–protein interactions in the networks of paralogous genes have great hierarchical relevance in terms of the overall regulatory networks, since the modified proteins commonly correspond to regulatory proteins that in turn regulate the transcription of many other genes or interact among these to form a cascade of transcriptional regulation, as in the cases of PhoP/PhoQ, RstA/RstB and PhoB/PhoR, two-component systems that respond to acidic conditions in *E. coli* (Ogasawara *et al.*, 2007).

The preceding examples demonstrating the regulatory role of the two-component system proteins provide evidence of the relevance of protein–protein interactions in the evolution of the regulatory networks of paralogous genes in *E. coli* and *B. subtilis*. It has been documented that the sensor domain of the histidine kinases is subject to a faster evolutionary process than that of the DNA-binding domain of the response-regulator proteins (Alm *et al.*, 2006). The plasticity of these elements in terms of their capacity to change their regulatory specificity as they respond to the adaptive process may include gene duplication and specialization (Commichau & Stulke, 2008). Furthermore, it has been proposed that two-component proteins acquired through duplication may undergo subsequent domain shuffling, permitting the development of novel genetic diversity. This may happen more quickly than that manifested by two-component systems acquired by horizontal gene transfer, as these proteins are more likely to retain their original functions (Alm *et al.*, 2006; Wanner, 1992; Yosef *et al.*, 2009). To sum up, massive duplication followed by shuffling, faster evolution in the sensor protein domain, and probably horizontal gene transfer, all have a significant influence on the evolution of protein–protein interactions and on the architecture of regulatory networks in bacteria.

RNA and RNA-binding proteins are important factors involved in shaping the regulatory networks of paralogous genes

The important role played by RNA molecules in gene regulation has been recently documented. In bacteria, RNA molecules can regulate either *in cis*, e.g. as riboswitches that are located in the 5′ untranslated leader region (Winkler & Breaker, 2005) or *in trans*, acting as diffusible products which act by base pairing with their target mRNAs, resulting in changes in their translation or stability (Gottesman *et al.*, 2006). We identified the participation of RNA elements in 28 and 52 groups of paralogous genes in *E. coli* and *B. subtilis*, respectively (Supplementary Tables S1 and S2), by means of computational analysis, using the co-variance models from the Rfam database (Gardner *et*

al., 2009). As an example of the RNA molecules recognized, Fig. 1(e) shows the regulation of the *E. coli rpoH* and *rpoS* paralogous genes (Group 374 of Supplementary Table S1), which code for the σ^H and σ^S sigma factors, respectively. The translation of *rpoH* is under the regulation of a *cis*-acting RNA element located in the coding sequence of the *rpoH* gene. At low temperatures, this thermo-regulator folds into a secondary structure which overlaps and sequesters the ribosome-binding site (RBS), inhibiting *rpoH* translation (Narberhaus *et al.*, 2006) (Fig. 1e). In contrast, the regulation of *rpoS* depends on a set of *trans*-acting small RNAs, in response to different kinds of stress factors, such as low temperature (DsrA), osmotic shock (RprA) and oxidative shock (OxyS). Interestingly, RprA and DsrA have a positive effect on *rpoS* translation by destabilizing a secondary structure that sequesters the RBS, whereas OxyS has a negative effect on translation, by an unidentified mechanism (Gottesman *et al.*, 2006) (Fig. 1e).

In general, the number of RNA molecules acting *in cis* as riboswitches on the regulatory networks of paralogous genes is greater in the Firmicute *B. subtilis* than in the Proteobacterium *E. coli*. Indeed, there are *B. subtilis* riboswitches that do not exhibit a counterpart in *E. coli*, which is the case for the T-box element, regulating different kinds of genes involved in amino acid biosynthesis, transport and charging onto their corresponding tRNAs (Gutierrez-Preciado *et al.*, 2005, 2009). On the other hand, the regulation of paralogous genes by *trans*-acting RNA elements is particularly important in *E. coli* and other Gram-negative bacteria, where it has been related to the response to different stress factors such as the aforementioned low temperature (DsrA), osmotic shock (RprA) and oxidative shock (OxyS), in addition to the stress caused by iron limitation (RyhB) and the accumulation of glucose phosphate (SgrR) (reviewed by Gottesman, 2005). In fact, there is an important bias of these small RNAs to regulate paralogous gene coding for outer-membrane proteins involved in stress response, previously described (see Groups 19, 58 and 278 of Supplementary Tables S1 and S3). Interestingly, some of these small RNAs have been generated by duplication, as is the case of *E. coli* OmrA RNA, which is accumulated in the late stationary phase, whereas its paralogous counterpart, OmrB, is transiently expressed in the early stationary phase (Guillier *et al.*, 2006). In higher organisms such as rice, it has been proposed that some pseudogenes may have acquired novel regulatory roles as antisense small RNAs, after being duplicated (Guo *et al.*, 2009). Duplication of regulatory RNA elements is not restricted to *trans*-acting RNAs, but may include *cis*-acting riboswitches. Such is the case with tandem T-boxes located in the regulatory regions of the tryptophan biosynthetic operons in *Bacillus anthracis*, *Bacillus cereus*, *Bacillus halodurans* and other Gram-positive bacteria. It has been proposed that these tandem ribositches may expand the range of regulation of their respective transcription units (Gutierrez-Preciado *et al.*, 2005).

RNA-binding proteins can also be included in the regulatory networks of genes generated by duplication. An example of this is found in the carbon catabolite repression system in *B. subtilis*, which makes possible the preferential utilization of easily metabolizable carbon sources, such as glucose. Fig. 1(f) shows the regulation of the paralogous genes *bglP* and *sacP*, which code for the permeases responsible for the transport of the β-glucoside salicin and sucrose, respectively (Group 36 of Supplementary Table S2). The transcriptional regulation of these genes is based on the paralogous RNA-binding proteins LicT and SacT, which can bind and disrupt potential ribonucleic acid anti-terminators located in the upstream regions of the *bglP* and *sacP* genes (Schilling *et al.*, 2006). As previously carried out for other regulatory elements, a statistical analysis was undertaken for the RNA regulatory interactions of *E. coli* and *B. subtilis*. Despite the participation of RNA elements in different paralogous groups, we only found 12 examples of an RNA element regulating two or more paralogous genes in *B. subtilis*. This value registers 11.6 SDs above the mean of their corresponding randomized regulatory networks.

The regulatory networks of paralogous genes are influenced by DNA supercoiling

Coordinate gene expression involves several hierarchical levels that comprise local control of individual genes or operons, multiple operons within a regulon and multiple regulons within a stimulon. It has been proposed that the highest level of hierarchical gene regulation is mediated by DNA supercoiling, which serves as a global regulator of gene expression (Hatfield & Benham, 2002). Although chromosomal DNA has slight negative supercoiling, the degree of supercoiling can be modulated by the action of DNA topoisomerases, in response to a wide variety of altered nutritional and environmental conditions, including temperature (Grau *et al.*, 1994), osmolarity (Cheung *et al.*, 2003; Grau *et al.*, 1994), oxidative stress (Weinstein-Fischer *et al.*, 2000), aerobic to anaerobic growth shift (Hsieh *et al.*, 1991) and growth stage in either *E. coli* (Cheung *et al.*, 2003) or *B. subtilis* (Bird *et al.*, 1992). DNA supercoiling adjusts the basal levels of expression of all genes whose transcription depends on promoters that are directly sensitive to supercoiling (Hatfield & Benham, 2002) or indirectly, acting in combination with some TFs or sigma factors. This is the case for some genes, where transcription by σ^{38} depends on the superhelical densities, occurring as a function of growth conditions (Bordes *et al.*, 2003). This fine-tuning of regulation by DNA supercoiling is also found to form part of the regulatory networks of paralogous genes. According to our data survey, in *E. coli* there are 15 groups of paralogous genes where regulation is directly modulated by DNA supercoiling. Interestingly, three of these groups correspond to DNA-binding transcriptional regulators which in turn exert a regulatory control on other genes within the network, thus expanding the regulatory response in a hierarchical manner. One of

these groups includes the OxyR regulator involved in the oxidative stress response (Group 4, See Supplementary Tables S1 and S3). In addition, we found groups of paralogous genes coding for transporters of amino acids and other types of metabolite, as well as genes coding for prophage proteins (Supplementary Tables S1 and S3). It is interesting that all the genes whose expression is reported to be modulated by DNA supercoiling belong to different paralogous groups; thus no statistical analysis was carried out to determine the probability of finding this regulatory element acting on two or more different gene targets.

Different kinds of regulatory elements simultaneously participate in the regulatory networks of paralogous genes

The response to different external and metabolic signals on the part of the regulatory networks of paralogous genes involves the integration of multiple kinds of regulatory elements, acting at different levels. An excellent example of this kind of regulatory integration is found in the *E. coli* paralogous porins OmpC and OmpF (Group 58 of Supplementary Table S3) in response to medium osmolarity, temperature and oxidative stress (Guillier *et al.*, 2006; Pratt *et al.*, 1996). Fig. 2 summarizes the different regulatory layers and types of regulatory elements of this system, which comprise: (i) regulation at transcription initiation by at least 11 TFs that act directly on *ompC* and *ompF*, or indirectly regulate the transcription of other regulatory elements of the system; (ii) regulation at translation initiation, accomplished by a set of small antisense RNAs; (iii) post-transcriptional control of the RNA chaperone StpA on *micF* RNA, by the induction of its degradation; and (iv) post-translational modification of protein activity, as is the case in OmpR activation, by means of its phosphorylation by the histidine kinase EnvZ.

Furthermore, Fig. 2 shows that a single regulatory element may act at different levels on more than one target gene. For example, H-NS may directly regulate *ompC* expression by blocking its transcription and indirectly regulate *ompF* translation by repressing the transcription of the *micF* antisense RNA. Interestingly, a specific TF may have a dual activity on its target genes and may act as a transcription activator or repressor, as a result of its interaction with other regulatory elements of the system. This is the case with OmpR, where the active state depends on its phosphorylation by EnvZ, in response to osmolarity. At a low concentration of OmpR-P (low osmolarity), transcription of *ompF* is activated. At high osmolarity, the concentration of OmpR-P increases and the protein binds to additional low-affinity sites, blocking the transcription of *ompF* (Pratt *et al.*, 1996).

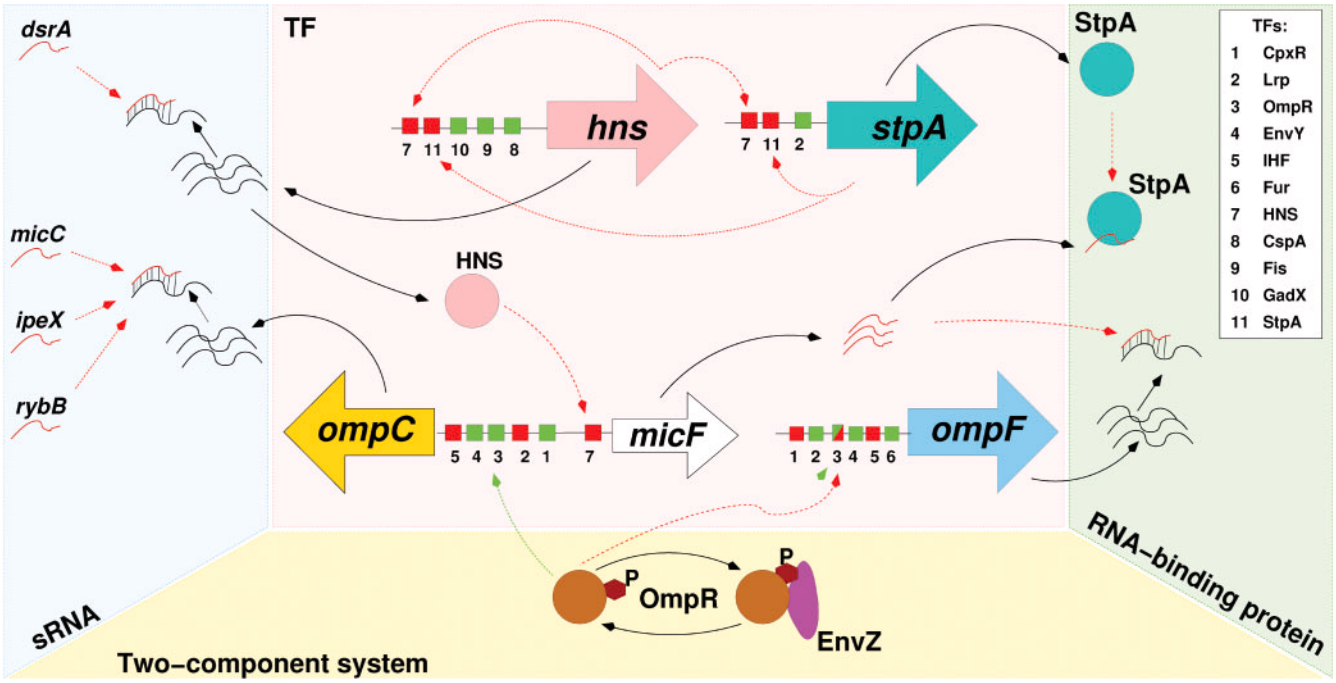


Fig. 2. Integrative example of a regulatory network of paralogous genes based on different kinds of regulatory elements. The regulatory network of the OmpC and OmpF paralogous porins is shown. Genes and their corresponding polypeptide products are depicted in the same colour. Dashed arrows represent regulatory interactions and are coloured according to their either positive (green) or negative (red) effect on expression. Small boxes in the intergenic sequences are coloured according to the regulatory effect of their corresponding TFs. The RNA gene products are represented by short, undulating lines and the protein products by ovals.

Concluding remarks

The compilation and analysis of regulatory elements in paralogous genes of model organisms has led us to extend the original model of the evolution of regulatory networks, which took only TFs into account, to a model of more general type, which includes other regulators such as small RNAs, riboswitches, RNA-binding proteins, sigma factors, protein–protein interactions and DNA supercoiling. Despite the fact that TFs are the most extensively used element in regulatory networks, the extended repertoire of other regulatory mechanisms has resulted in a significant increase in the versatility of the network, accurately modulating the organism's gene expression. Our analysis of the model organisms *E. coli* and *B. subtilis* is consistent with previous observations (Conant & Wolfe, 2008) regarding the relevance of the regulatory networks of paralogous genes in the process of adaptation of an organism in response to changing environments.

Acknowledgements

We thank Cei Abreu-Goodger and Zuemy Rodriguez-Escamilla for their critical reading of the manuscript, Ricardo Ciria for computer support and Shirley Ainsworth for bibliographical assistance. This work was supported by CONACyT grants 60127-Q and SALUD-2007-C01-68992 to E. M. and DGAPA grant IN217508 to E. P.-R.

References

Alm, E., Huang, K. & Arkin, A. (2006). The evolution of two-component systems in bacteria reveals different strategies for niche adaptation. *PLOS Comput Biol* **2**, e143.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M. & Teichmann, S. A. (2004). Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14**, 283–291.

Babu, M. M., Teichmann, S. A. & Aravind, L. (2006). Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol* **358**, 614–633.

Balaji, S., Babu, M. M. & Aravind, L. (2007). Interplay between network structures, regulatory modes and sensing mechanisms of transcription factors in the transcriptional regulatory network of *E. coli*. *J Mol Biol* **372**, 1108–1122.

Bausch, C., Ramsey, M. & Conway, T. (2004). Transcriptional organization and regulation of the L-idonic acid pathway (GntII system) in *Escherichia coli*. *J Bacteriol* **186**, 1388–1397.

Bird, T., Burbulys, D., Wu, J. J., Strauch, M. A., Hoch, J. A. & Spiegelman, G. B. (1992). The effect of supercoiling on the in vitro transcription of the *spoIIA* operon from *Bacillus subtilis*. *Biochimie* **74**, 627–634.

Bordes, P., Conter, A., Morales, V., Bouvier, J., Kolb, A. & Gutierrez, C. (2003). DNA supercoiling contributes to disconnect σ^S accumulation from σ^S-dependent transcription in *Escherichia coli*. *Mol Microbiol* **48**, 561–571.

Cheung, K. J., Badarinarayana, V., Selinger, D. W., Janse, D. & Church, G. M. (2003). A microarray-based antibiotic screen identifies a regulatory role for supercoiling in the osmotic stress response of *Escherichia coli*. *Genome Res* **13**, 206–215.

Chilcott, G. S. & Hughes, K. T. (2000). Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar typhimurium and *Escherichia coli*. *Microbiol Mol Biol Rev* **64**, 694–708.

Commichau, F. M. & Stulke, J. (2008). Trigger enzymes: bifunctional proteins active in metabolism and in controlling gene expression. *Mol Microbiol* **67**, 692–702.

Conant, G. C. & Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* **9**, 938–950.

Evangelisti, A. M. & Wagner, A. (2004). Molecular evolution in the yeast transcriptional regulation network. *J Exp Zool B Mol Dev Evol* **302**, 392–411.

Ferris, H. U. & Minamino, T. (2006). Flipping the switch: bringing order to flagellar assembly. *Trends Microbiol* **14**, 519–526.

Fondi, M., Brilli, M., Emiliani, G., Paffetti, D. & Fani, R. (2007). The primordial metabolism: an ancestral interconnection between leucine, arginine, and lysine biosynthesis. *BMC Evol Biol* **7** (Suppl 2), S3.

Gama-Castro, S., Jimenez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Penaloza-Spinola, M. I., Contreras-Moreira, B., Segura-Salazar, J., Muniz-Rascado, L., Martinez-Flores, I. & other authors (2008). RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res* **36**, D120–D124.

Gardner, P. P., Daub, J., Tate, J. G., Nawrocki, E. P., Kolbe, D. L., Lindgreen, S., Wilkinson, A. C., Finn, R. D., Griffiths-Jones, S. & other authors (2009). Rfam: updates to the RNA families database. *Nucleic Acids Res* **37**, D136–D140.

Gelfand, M. S. (2006). Evolution of transcriptional regulatory networks in microbial genomes. *Curr Opin Struct Biol* **16**, 420–429.

Gottesman, S. (2005). Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends Genet* **21**, 399–404.

Gottesman, S., McCullen, C. A., Guillier, M., Vanderpool, C. K., Majdalani, N., Benhammou, J., Thompson, K. M., FitzGerald, P. C., Sowa, N. A. & FitzGerald, D. J. (2006). Small RNA regulators and the bacterial response to stress. *Cold Spring Harb Symp Quant Biol* **71**, 1–11.

Grau, R., Gardiol, D., Glikin, G. C. & de Mendoza, D. (1994). DNA supercoiling and thermal regulation of unsaturated fatty acid synthesis in *Bacillus subtilis*. *Mol Microbiol* **11**, 933–941.

Gu, Z., Nicolae, D., Lu, H. H. & Li, W. H. (2002). Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet* **18**, 609–613.

Gu, X., Zhang, Z. & Huang, W. (2005). Rapid evolution of expression and regulatory divergences after yeast gene duplication. *Proc Natl Acad Sci U S A* **102**, 707–712.

Guillier, M., Gottesman, S. & Storz, G. (2006). Modulating the outer membrane with small RNAs. *Genes Dev* **20**, 2338–2348.

Guo, X., Zhang, Z., Gerstein, M. B. & Zheng, D. (2009). Small RNAs originated from pseudogenes: cis- or trans-acting? *PLOS Comput Biol* **5**, e1000449.

Gutierrez-Preciado, A., Jensen, R. A., Yanofsky, C. & Merino, E. (2005). New insights into regulation of the tryptophan biosynthetic operon in Gram-positive bacteria. *Trends Genet* **21**, 432–436.

Gutierrez-Preciado, A., Henkin, T. M., Grundy, F. J., Yanofsky, C. & Merino, E. (2009). Biochemical features and functional implications of the RNA-based T-box regulatory mechanism. *Microbiol Mol Biol Rev* **73**, 36–61.

Hatfield, G. W. & Benham, C. J. (2002). DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu Rev Genet* **36**, 175–203.

Hecker, M., Pane-Farre, J. & Volker, U. (2007). SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. *Annu Rev Microbiol* **61**, 215–236.

Hengge-Aronis, R. (2002). Signal transduction and regulatory mechanisms involved in control of the σ^S (RpoS) subunit of RNA polymerase. *Microbiol Mol Biol Rev* **66**, 373–395 (table).

Hsieh, L. S., Burger, R. M. & Drlica, K. (1991). Bacterial DNA supercoiling and [ATP]/[ADP]. Changes associated with a transition to anaerobic growth. *J Mol Biol* **219**, 443–450.

Izu, H., Adachi, O. & Yamada, M. (1997). Gene organization and transcriptional regulation of the *gntRKU* operon involved in gluconate uptake and catabolism of *Escherichia coli*. *J Mol Biol* **267**, 778–793.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S. & other authors (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res* **36**, D480–D484.

Kashtan, N. & Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci U S A* **102**, 13773–13778.

Kroos, L. & Yu, Y. T. (2000). Regulation of sigma factor activity during *Bacillus subtilis* development. *Curr Opin Microbiol* **3**, 553–560.

Li, W. H., Yang, J. & Gu, X. (2005). Expression divergence between duplicate genes. *Trends Genet* **21**, 602–607.

Lioliou, E. E., Mimitou, E. P., Grigoroudis, A. I., Panagiotidis, C. H., Panagiotidis, C. A. & Kyriakidis, D. A. (2005). Phosphorylation activity of the response regulator of the two-component signal transduction system AtoS-AtoC in *E. coli*. *Biochim Biophys Acta* **1725**, 257–268.

Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A. & Gerstein, M. (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431**, 308–312.

Lynch, M. & Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155.

Martinez-Antonio, A. & Collado-Vides, J. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol* **6**, 482–489.

Moreno-Campuzano, S., Janga, S. C. & Perez-Rueda, E. (2006). Identification and analysis of DNA-binding transcription factors in *Bacillus subtilis* and other Firmicutes – a genomic approach. *BMC Genomics* **7**, 147.

Narberhaus, F., Waldminghaus, T. & Chowdhury, S. (2006). RNA thermometers. *FEMS Microbiol Rev* **30**, 3–16.

Ninfa, A. J. & Magasanik, B. (1986). Covalent modification of the *glnG* product, NRI, by the *glnL* product, NRII, regulates the transcription of the *glnALG* operon in *Escherichia coli*. *Proc Natl Acad Sci U S A* **83**, 5909–5913.

Ogasawara, H., Hasegawa, A., Kanda, E., Miki, T., Yamamoto, K. & Ishihama, A. (2007). Genomic SELEX search for target promoters under the control of the PhoQP-RstBA signal relay cascade. *J Bacteriol* **189**, 4791–4799.

Pragai, Z., Allenby, N. E., O'Connor, N., Dubrac, S., Rapoport, G., Msadek, T. & Harwood, C. R. (2004). Transcriptional regulation of the *phoPR* operon in *Bacillus subtilis*. *J Bacteriol* **186**, 1182–1190.

Pratt, L. A., Hsing, W., Gibson, K. E. & Silhavy, T. J. (1996). From acids to *osmZ*: multiple factors influence synthesis of the OmpF and OmpC porins in *Escherichia coli*. *Mol Microbiol* **20**, 911–917.

Reitzer, L. & Schneider, B. L. (2001). Metabolic context and possible physiological themes of σ^{54}-dependent genes in *Escherichia coli*. *Microbiol Mol Biol Rev* **65**, 422–444.

Schilling, O., Herzberg, C., Hertrich, T., Vorsmann, H., Jessen, D., Hubner, S., Titgemeyer, F. & Stulke, J. (2006). Keeping signals straight in transcription regulation: specificity determinants for the interaction of a family of conserved bacterial RNA-protein couples. *Nucleic Acids Res* **34**, 6102–6115.

Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* **31**, 64–68.

Sierro, N., Makita, Y., de Hoon, M. & Nakai, K. (2008). DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Res* **36**, D93–D96.

Teichmann, S. A. & Babu, M. M. (2004). Gene regulatory network growth by duplication. *Nat Genet* **36**, 492–496.

Typas, A., Becker, G. & Hengge, R. (2007). The molecular basis of selective promoter activation by the σ^S subunit of RNA polymerase. *Mol Microbiol* **63**, 1296–1306.

Wanner, B. L. (1992). Is cross regulation by phosphorylation of two-component response regulator proteins important in bacteria? *J Bacteriol* **174**, 2053–2058.

Weinstein-Fischer, D., Elgrably-Weiss, M. & Altuvia, S. (2000). *Escherichia coli* response to hydrogen peroxide: a role for DNA supercoiling, topoisomerase I and Fis. *Mol Microbiol* **35**, 1413–1420.

Winkler, W. C. & Breaker, R. R. (2005). Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* **59**, 487–517.

Yamamoto, K. & Ishihama, A. (2005a). Transcriptional response of *Escherichia coli* to external copper. *Mol Microbiol* **56**, 215–227.

Yamamoto, K. & Ishihama, A. (2005b). Transcriptional response of *Escherichia coli* to external zinc. *J Bacteriol* **187**, 6333–6340.

Yosef, N., Kupiec, M., Ruppin, E. & Sharan, R. (2009). A complex-centric view of protein network evolution. *Nucleic Acids Res* **37**, e88.

Zhang, J., Rosenberg, H. F. & Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A* **95**, 3708–3713.

Zimmer, D. P., Soupene, E., Lee, H. L., Wendisch, V. F., Khodursky, A. B., Peter, B. J., Bender, R. A. & Kustu, S. (2000). Nitrogen regulatory protein C-controlled genes of *Escherichia coli*: scavenging as a defense against nitrogen limitation. *Proc Natl Acad Sci U S A* **97**, 14674–14679.