Science Progress (2012), 95(3), 315-329
doi: 10.3184/003685012X13420097673409

The repertoire of DNA-binding transcription factors in prokaryotes: functional and evolutionary lessons

ERNESTO PEREZ-RUEDA AND MARIO ALBERTO MARTINEZ-NUÑEZ

ABSTRACT

The capabilities of organisms to contend with environmental changes depend on their genes and their ability to regulate their expression. DNA-binding transcription factors (TFs) play a central role in this process, because they regulate gene expression positively and/or negatively, depending on the operator context and ligand-binding status. In this review, we summarise recent findings regarding the function and evolution of TFs in prokaryotes. We consider the abundance of TFs in bacteria and archaea, the role of DNA-binding domains and their partner domains, and the effects of duplication events in the evolution of regulatory networks. Finally, a comprehensive picture for how regulatory networks have evolved in prokaryotes is provided.

Keywords: *transcription factors, regulatory networks, bacteria, archaea, genomics, DNA-binding domain, partner domain*

Ernesto Perez-Rueda has been a professor at Universidad Nacional Autónoma de México since 2004. His research focuses on the understanding of the functional role and evolution of DNA-binding transcription factors in bacteria and archaea. He has published several international works on these topics, such as the repertoire of DNA-binding transcription factors in the bacteria *Escherichia coli K12* and *Bacillus subtilis 168*. He may be contacted at Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62100, Mexico. E-mail: erueda@ibt.unam.mx

Mario Alberto Martínez-Núñez has been a post-doctor at the Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México since 2010. His work focuses on the analysis of divergence of regulation among paralogous genes in the bacteria *Escherichia coli K12* and *Bacillus subtilis 168*. He has published works on the evolution of transcriptional regulatory networks in bacteria.

Introduction

Adaptive responses associated with environmental changes include the modification of the genetic program and, as a consequence, changes in metabolism. In all organisms, it is well known that gene expression is modulated predominantly at the level of transcription initiation through DNA-binding transcription factors (TFs), which provide the ability to express different genes under different metabolic stimuli or growth conditions. In this sense, TFs repress or activate gene expression by blocking or allowing the access of the RNA polymerase (RNAP) to the promoter, depending on the operator context and ligand-binding status[1]. Due to the crucial role that TFs have in coordinating the gene expression, they have been widely studied in different ways, including mutation analysis and the elucidation of numerous three-dimensional structures. On a genomic scale, three organisms have been considered as the model to analyse in detail the repertoires of TFs, *Escherichia coli* K-12[2], *Bacillus subtilis* 168[3], and *Corynebacterium glutamicum*[4]. These repertoires together with the structural assignations have allowed us to evaluate the distribution and abundance of this class of proteins in different organisms[5-7]. From these studies, it has been observed that the number of TFs increases from a few in the archaea *Nanoarchaeum equitans* Kin4-M and *Ignococcus hospitalis* to hundreds in the bacteria *Bacillus subtilis* and *E. coli* K-12[6,8]. This increment correlates with the hypothesis of genome maturation, which proposes that a greater number of regulatory elements are necessary to regulate a greater number of genes. Consequently, the number of genetic circuits in the regulatory networks that arise also increases as well as the interconnections between them[9]. Therefore, minor changes in single genes may propagate along such networks and may produce, in the end, quite drastic effects on gene expression in response to internal and external stimuli. In view of the importance of TFs in gene regulation, in this work we summarise some of the most recent insights from studies on the function and evolution of this class of proteins in prokaryotes and their probable implications in the evolution of regulatory networks. We break the subject into sections, covering the abundance of TFs, the role of DBDs and their partner domains (PaDos), and the effects of duplication events in the context of a regulatory network. We finish with some conjectures that attempt to provide a comprehensive picture of how regulatory networks have evolved in prokaryotes and the role of TFs in this organisation.

TFs can regulate via different combinatorial mechanisms

TFs constitute a class of proteins whose sphere of action is highly flexible, not only in sensing diverse environmental and endogenous stimuli but also in exploring various combinatorial mechanisms that can differentially regulate gene expression. In this regard, TFs can associate in a combinatorial fashion to regulate differentially as a consequence of metabolic signals. For instance, the *nirBCD-cysG* operon, which encodes a nitrite reductase involved in the reduction of nitrite to ammonia in *E. coli* K-12, is regulated by up to eight different TFs responsible for various cellular responses, including FruR (fructose assimilation regulatory protein), NarL and NarP (nitrate/nitrite regulators), and Fnr (fumarate nitrate reduction/regulator of anaerobic respiration)[10]. Another example is cell differentiation in the bacterium *Myxococcus xanthus*, in which two operons, *fmgA* (C-signal-dependent protein) and *fmgBC* (reductase and oxidase components), are regulated by the combinatorial effect of FruA and MrpC2 TFs. These TFs bind cooperatively in the promoters, although the arrangements of binding sites differ, *i.e.*, whereas FruA binds upstream of MrpC2 in the *fmgA* promoter region, MrpC2 binds upstream of FruA in the *fmgBC* promoter region[11,12]. In addition, the diversity of sequences recognised by TFs is enormous and can occur anywhere from a few bases downstream of the promoter zone to up to hundreds of bases upstream

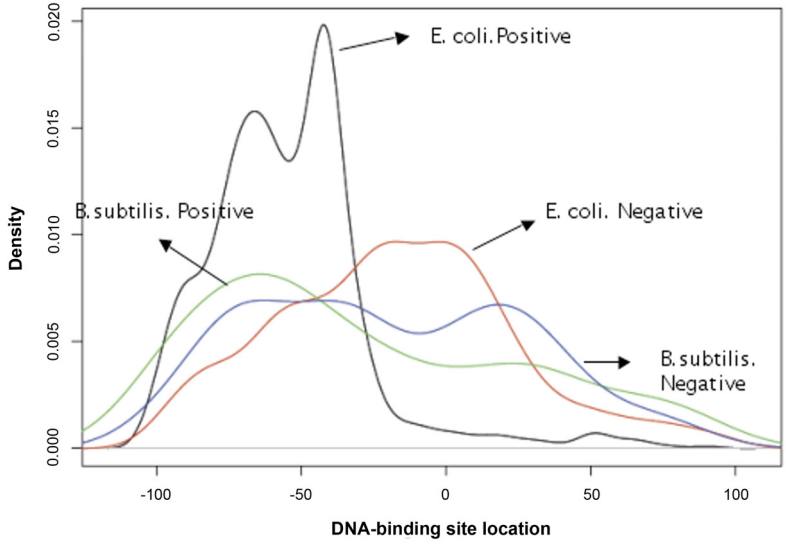

Figure 1 *Distribution of TF DNA-binding sites in* E. coli *and* B. subtilis*. Binding sites are preferentially located between the –100 and –20 positions relative to the transcription start for positive regulation and around the +1 position for repression. In this plot, 0 represents the +1 site, and density corresponds to the proportion of DNA-binding sites located in a particular position.*

of the transcription start site, enhancing or repressing gene expression (Figure 1). In this regard, four main modes of repression associated with TFs and the location of their DNA-binding sites in bacteria have been described: (a) repression by steric hindrance, where the TF binds between or over the core promoter elements, (b) repression by blocking transcription elongation, where the TF binds at the start of the coding region, (c) repression by DNA looping, where the DNA-binding sites are upstream and downstream of the promoter and two monomers of the same TF bend the DNA, and (d) repression by modulation of an activator. In contrast, four main modes of activation mediated by TFs have also been proposed: (a) class I activation, where a TF binds upstream of the promoter and interacts with the alpha-subunit of RNAP; (b) class II, in which a TF binds the DNA upstream of the core promoter and promotes σ factor binding; (c) activation mediated by conformational changes, where a TF binds to the promoter to enable it to be bound to the RNAP, often by twisting the DNA; and (d) activation by modulation of a repressor, alleviating the repression effect (this is also called antirepression)[1]. Furthermore, the interplay of TFs with σ factors in bacteria enhances the diversification of regulatory mechanisms, such as CcpA of *B. subtilis*, a global regulatory protein involved in catabolite repression, that may act as a positive regulator of genes involved in excretion of carbon excess and that can associate with three different σ factors (σ^A, σ^L, and σ^E) and with more than 10 different TFs[13]. In archaea, similar regulatory principles can be suggested because their similarity to bacterial TFs and operonic organisation[14].

The diversity of TFs is constrained to a few DNA-binding domains

The structures of more than 30 prokaryotic DNA-binding TFs have now been determined, and hundreds of amino acid sequences are known for many more. This kind of information together with the genomic data have allowed us to evaluate the domain organisation of this class of proteins in different organisms In this regard, Ulrich *et al*. [15] described that up to 84% of the DNA-binding domains (DBDs) in one-component TFs comprise a DNA-binding helix–turn–helix (HTH); whereas Charoensawan *et al*. and Madan Babu and Teichmann[2,7] suggest that the most abundant DBD in prokaryotes is the winged helix DNA-binding domain (wHTH), *i.e.* around 45% of the total set of TFs contains this domain, followed by the homeodomain-like (~26%) and lambda repressor-like DNA-binding domain (lHTH) (~10%) (see Table 1). In addition, these domains are widely distributed among the bacteria and archaea organisms, suggesting

Table 1 *Abundance of DBDs of TFs in bacteria and archaea*

Superfamily[a]	Proportion[b]	Families[c]	TFs[c]
Winged helix DNA-binding domain	44.18	CRP, LysR, GntR, AsnC, MarR	Crp of *E. coli*, Lrs14 and SSO1352 of *Sulfolobus solfataricus*
Homeodomain-like	26.30	AraC/XylS, TetR/AcrR, Fis	NtrC of *E. coli*, LmrA of *B. subtilis*, TM_1030 of *Thermotoga maritima*
Lambda repressor-like DNA-binding domains	10.55	GalR/LacI	CcpA of *B. subtilis*, GalR of *E. coli*
C-terminal (C-term) effector domain of the bipartite response regulators	9.68	OmpR, Spo0A, GerE-like (LuxR/UhpA family of transcriptional regulators)	PhoB and NarL of *E. coli*
Putative DNA-binding domain	3.02	MerR	CueR of *E. coli*, GlnR of *B. subtilis*
Nucleic acid-binding proteins	1.86	Cold shock DNA-binding domain-like	CspA of *E. coli*
Putative transcriptional regulator TM1602, C-terminal domain	1.75	HTH_11	CodY and BirA of *B. subtilis*
IHF-like DNA-binding proteins	0.72	IHF	IHF of *E. coli*
AbrB/MazE/MraZ-like	0.71	AbrB N-terminal domain-like	Transcription regulator AbrB, of *B. subtilis*
Ribbon-helix-helix	0.19	Met repressor, MetJ (MetR), Arc/Mnt-like phage repressors, CopG	NikR, MetJ of *E. coli*
TrpR-like	0.19	Trp repressor, TrpR	TrpR of *E. coli* K-12
KorB DNA-binding domain-like	0.10	KorB DNA-binding domain-like	Transcriptional repressor protein KorB of *E. coli*
Flagellar transcriptional activator FlhD	0.08	Flagellar transcriptional activator FlhD	FlhD of *E. coli*
H-NS histone-like proteins	0.67	HN-S	H-NS of *E.coli*

[a]DBD Superfamily name.
[b]Proportion of proteins containing a domain (values are in percentage).
[c]Families and TFs more representatives associated to each superfamily.
DBDs were retrieved from DBD, Supfam and PFAM databases. Superfamilies are sorted according their abundance.

an ancient origin as was previously noted by Aravind and Koonin [16]. In this regard, the authors suggest that the DBDs associated with TFs are among the most ancient domains described so far, derived from a relatively small set of folds. Alternative DBDs, such as, ribbon–helix–helix, IHF-like DNA-binding domain, PhoU-like domain, nucleic acid-binding domain associated to cold shock proteins (CSD) were also identified, although in lower proportions[2,7,8]. Indeed, some of these DBDs have been also identified as being constrained to specific phyla, such as the ribbon–helix–helix and TrpR being identified almost exclusively in gammaproteobacteria, whereas other DBDs, such as the nucleic acid-binding domain associated with cold shock proteins might have been acquired after the prokaryotes and eukaryotes split by lateral gene transfer from the eukaryotes[8,17].

An important question that remains to be explored concerns the diversity of DBDs associated to TFs and their evolutionary and functional implications. In this context, Itzkovitz *et al.*[18] suggest that the apparent limit on the diversity of DNA-binding structures associated with TFs correlates with the number of DNA bases effectively recognised, minimising the cross-binding errors between TFs. Therefore, an evolutionary scenario for the TFs is proposed, in which "simple" organisms, which require few TFs, employ certain DBD proteins, such as the helix–turn–helix. When these DBDs reach their upper bounds, new DNA-binding structures are needed, and organisms shift their TF usage to novel structures with more degrees of freedom and higher maximal numbers, such as occurs with the use of the C_2H_2 zinc–finger or helix–loop–helix TFs in eukaryotes [18]. In this sense, alternative DNA-binding structures could increasing the diversity of TFs to regulate gene expression in bacteria and archaea, such as the RelE-like structure identified in the regulator RelE of *E. coli* K-12, associated to the toxin–antitoxin systems[19].

Abundance of TFs correlates with genome size in prokaryotes

Recent studies have shown that the evolutionary events associated with regulatory proteins, such as their expansion and contraction, contribute significantly in shaping the gene content and genome size of the different lineages of prokaryotes[8]. Based on comparative genomics, a positive correlation between the growth of TF families and the variations in bacterial genome sizes has been described[20,21], with greater overrepresentation of TFs in large genomes than in small genomes (Figure 2). In this regard, in bacteria genomes the TF expansion follows a power law increase with an exponent close to 2, which infers a quadratic increase. In contrast, in archaea, where organisms contain a low

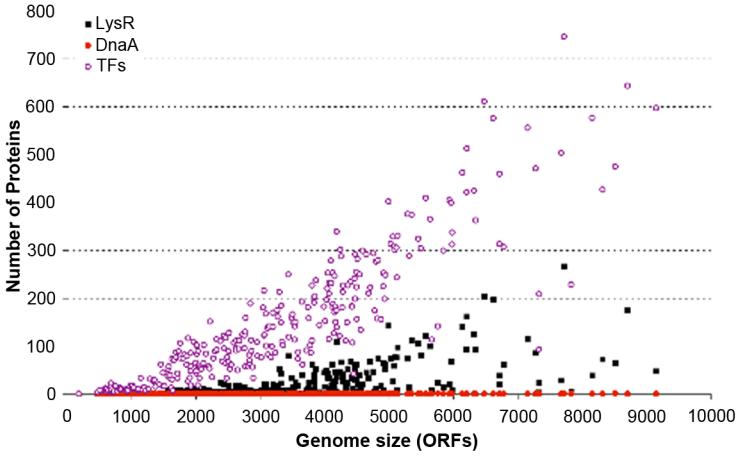

Figure 2 *Distribution of the number of TFs in prokaryotes as a function of genome size. Genomes are presented on the X-axis as the number of ORFs. Abundance of TFs in each genome is shown on the Y-axis (each dot corresponds to one genome).The large family LysR and the small family DnaA are shown in comparison to the total TFs.*

proportion of TFs behaving as intracellular pathogens or opportunistic pathogens, even though they exhibit genome sizes similar to free-living bacteria, a linear correlation has been reported[5,6]. A plausible hypothesis is that the abundance of TFs increases with an increase in an organism's complexity[20,21] and the subsequent need to coordinate and couple the expression of most genes and cellular functions. In addition, a possible functional relationship between TFs and prokaryote lifestyles could also influence the observed trend. Thus, to understand how the complexity of gene regulation depends on the number of TFs as a function of increasing genome size and how they are associated with the organism's lifestyle, the bacterial organisms have been classified into four global lifestyle classes[22]. These classes include extremophiles, intracellular bacteria, pathogens, and free-living bacteria. From this analysis, it was suggested that bacterial and archaeal complexity influences the repertoire of TFs, as these proteins increase in relation to the genome size in all lifestyle groups. These results suggest that a few regulatory elements identified in small genomes would compensate for the regulation of the entire genome, with an increase in the number of DNA-binding sites per element, in contrast to the large number of elements identified in large genomes, which control a smaller proportion of DNA-binding sites on average. In addition, a larger proportion of genes in small genomes are organised in operons, simplifying the transcriptional machinery necessary for gene expression, in contrast to large genomes, which have reduced numbers of genes in operons[23], which would also influence the proportion of TFs

in those organisms, suggesting that complex lifestyles require a higher proportion of TFs to better orchestrate a response to changing conditions. Therefore, the TF repertoires observed in bacteria and archaea correlate with the genome size and with the lifestyles and are the result of two main evolutionary mechanisms, the loss of genes by mutation and deletion and the acquisition of new genes either by horizontal gene transfer or by gene duplication[22]. Recently, an appealing hypothesis described in archaeal TFs suggested that the formation of modular complexes, as occurs in eukaryotic regulatory systems, contributes to compensate for the apparent deficit of TFs, increasing the complexity of regulation in prokaryotes[5,6].

Abundance of TFs does not correlate with diversity of families, and large families are not the most widely distributed

TFs can be grouped into families that reflect functional similarities and common evolutionary histories. In recent studies, it has been proposed that the diversity and abundance of TF families contribute to the regulatory plasticity[24,25], with a reduced diversity of families in small genomes, especially in pathogens and free-living organisms, and an increasing proportion in larger ones. The diversity of families reaches a maximum in genomes with around 5,000 genes. However, the higher number of TFs in larger genomes does not necessarily imply diversity of families beyond this plateau, but instead implies an increase in the size of some families of TFs. Indeed, the average number of TFs per family increases linearly, with a few families of TFs expanding disproportionately[24,25]. These families comprise LysR and TetR, which represent about 25% of the total set of TFs in prokaryotes (see Figure 2). Members of these two families increase abruptly in larger genomes and coincide with the plateauing of the diversity of families in bacterial and archaeal genomes. Another feature associated with large families is that they are not widely distributed among prokaryotes despite their role in controlling important processes, such as cell–cell communication (LuxR), sensing, uptake, and metabolism of external food sources (GntR), and resistance to multiple compounds (TetR), among other families. Alternatively, families with few copies per genome, such as DnaA, BirA, and Fur, which have been proposed to be essential under standard growth conditions in *E. coli* and in maintaining DNA, biotin biosynthesis, and metal homeostasis, uptake, respectively[10], might be considered universal in bacteria, because they have been identified in at least 80% of the genomes, suggesting gene loss events in organisms in which they are absent.

In summary, small families widely distributed among bacteria might be

related to ancestral functions beyond transcriptional regulation, such as DNA organisation or nucleoid integrity. In addition, these small families may contain global regulators, as has been found for the leucine repressor protein from the AsnC family in *E. coli* and CcpA (LacI family) in *B. subtilis*. In contrast, large families might be associated with specialisation of emergent processes, such as those involved in quorum sensing, as seen with members of the LuxR family. Indeed, the evolution of this mechanism in bacteria has been proposed to be one of the early steps in the development of multicellularity[26] and may be correlated with bacterial specialisation.

Contribution of partner domains to the diversity of the transcriptional machinery

An important aspect of TFs is their modular organisation in terms of structural domains. In this regard, information about the domain structures associated to TFs in *E. coli* K-12 suggest that the DBDs generally occur in combination with other domains, where there is a predominance of two-domain proteins (~75%) followed by three-domain proteins (~12%), single-domain proteins (~10%) and, finally, four-domain proteins (~3%)[2]. A similar distribution has been described in the repertoire of TFs in *B. subtilis*, *i.e.* a predominance of two-domain proteins[3]. In contrast, in archaeal genomes around three quarters (~75%) of their TFs have been identified as single-domain proteins, whereas multidomain TFs are distributed in low proportions, contrasting dramatically with the domain organisation of TFs in bacteria[5,6]. Thus, the domain organisation provides important clues about how signals, as small metabolites and covalent modifications, translate the environmental conditions into the response(s) by activating or deactivating the TFs, which, in turn, induce or repress the transcription of specific genes or operons. In this sense, the partner domains or PaDos play a fundamental role linking environmental conditions and mRNA synthesis. In general, these domains are associated with diverse functions, such as allosteric regulation of TFs across binding to a wide variety of functional compounds, in protein–protein interactions, and with enzymatic properties. Despite the importance of these domains, few analyses have been performed to explore them, such as in the GntR family, for which four subfamilies that correlate with the functions of the regulated genes have been identified [27]. In this regard, Rivera-Gomez *et al*.[28] evaluated the contributions of the domain organisation in the total set of proteins identified as TFs with a winged (wHTH) domain in bacteria, the most abundant structure in regulatory proteins. From this study, different groups were identified based on domain architecture, such as those TFs with more than one structural domain, *i.e.*, the DBD

and diverse PaDos (57%) and TFs with only the DBD, *i.e.*, monodomain proteins (43%). Indeed, a high diversity of PaDos identified in the whole collection of wHTH TFs was found, such as periplasmic-binding protein-like II, cAMP-binding domain-like, GAF domain-like, and LexA/signal peptidase domains. These proteins domains are also associated with the largest families, such as LysR, and they are phylogenetically widely distributed, suggesting that these PaDos are very evolutionary successful in all the bacteria. Another group of PaDos corresponds to domains widely distributed in bacteria, except that they are absent in parasites, symbionts and, in general, in small genomes, suggesting probable gene loss events, such as dimeric α- and β-barrel, PLP-dependent transferases and iron domains. Additional domains have been identified as being associated with specific divisions, such as the MOP-like, S-adenosyl-L-methionine-dependent methyltransferases and acyl-CoA N-acyltransferases, which have been mainly identified in the proteobacteria and acidobacteria divisions. Similar results have been observed with additional DBDs, suggesting a similar pattern distribution and specific associations between the DBDs and their PaDos. In summary, the diversity in the repertoire of regulatory proteins seems to be influenced by the organisation and combination with the PaDos and allows classification of the families into three groups (Figure 3): (a) monodomain families, where the

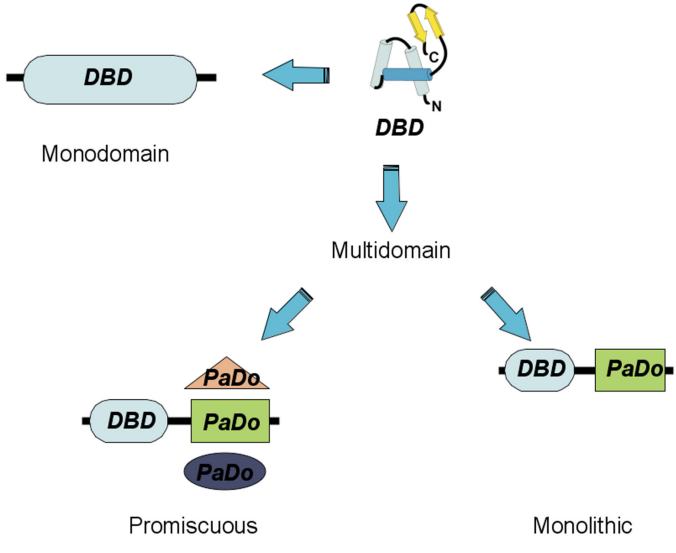

Figure 3 *Domain diversity associated with TFs. In monodomain TFs, the DBD covers most of the sequence, such as that found with the Fur family. In monolithic families, where at least 80% of the members exhibit a predominant PaDo associated with the DBD, and promiscuous families, such as GntR, diverse domains are associated with the DBD.*

multimerisation and ligand-binding sites are included in the DBD, such as occurs in archaeal TFs; (b) promiscuous families, those TFs with a large diversity of domains, such as occurs in the GntR family which contains a large diversity of PaDos; and (c) monolithic families, where the DBD has undergone a similar evolutionary process as the PaDos with few recombination events, as found in the LysR family. Therefore, the domain organisation associated with the DBD and PaDos and the family abundance associated with duplication events would contribute to increase the regulatory plasticity in prokaryotes, among other mechanisms described in this work. Finally, we have mainly centred the discussion of partner domains in one-component TFs because they are the most abundant regulatory proteins described so far in bacteria and archaea cellular domains[13]; however, two-component proteins not discussed here represent a large and important class of regulatory proteins that also contribute to the regulatory plasticity previously described.

Role of TFs in the evolution of regulatory networks

The variation in transcriptional regulatory mechanisms plays an important role during the origin and adaptation of species. Simple modifications within the upstream regulation region of a TF can explain both minor and major changes between species without involving any disruptions of gene structure. Mutations in the "consensus" regulatory region will not only cause quantitative changes in the binding affinity but also lead to complete dissociation or even substitution to rewire novel TFs [29]. In this regard, regulatory interactions can be conceptualised as transcriptional regulatory networks (TRNs), where nodes correspond to genes or TFs and edges represent the effects of the regulatory interactions, *i.e.*, activation or repression. From this perspective, global regulators have been identified as nodes highly connected, while the majority of TFs or local regulators exhibit a low number of connections. This hierarchical architecture provides robustness against fluctuations, as mutations. TFs can also form regulatory motifs, which contribute to the TRN circuitry and couple gene expression to environmental signals, such as the feed-forward loop, allowing cells to act as filters to transient signals and to maintain the expression of a gene at an “adequate” concentration[30].

From this perspective, the diversity of TRNs seems to increase with the evolution of regulatory regions and duplication of their elements (TFs and interactions)[31,32]. Duplication events of TFs allow a more versatile adaptation of the functional divergence gained from the duplication of structural genes. Initially, the duplicated TF recognises the same DNA motif as the parental regulator, and later it differentiates to interact with new DNA-binding sites, becoming a novel regulator responding to an

alternative environment, such as the duplicated metalloregulators Fur, Zur, and PerR, which are involved in iron, zinc, and iron-dependent peroxide sensing, respectively[10]. In general, the loss and gain of interactions between TFs and their regulated genes contribute significantly to the divergence of regulation in TRNs (Figure 4). Recently, Martinez-Nuñez *et al.*[33] described the evolutionary plasticity of the regulatory networks,

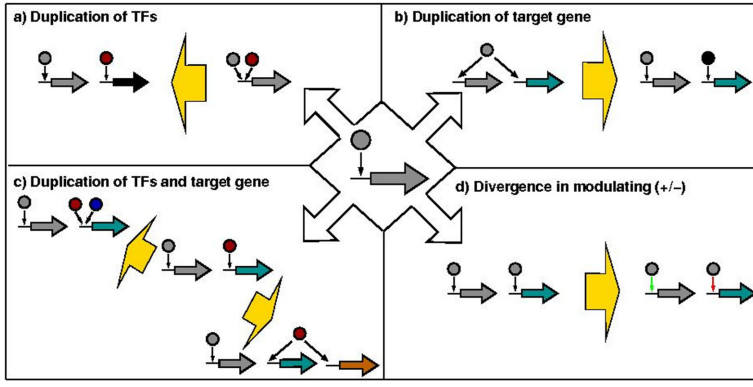

Figure 4 *Model of the evolution of TRNs. The model proposes that the loss and gain of regulatory interactions may occur following the duplication of either a TF (a) or a target gene (b), or following the duplication of both a TF as well as a target gene (c). Extension of the Teichmann and Babu model[34], which proposed the acquisition of a new regulatory mechanism through the divergence in the modulation (positive or negative) exerted on the target gene is also shown (d).*

which is not only the result of the duplication of TF interactions within a regulatory network, as suggested Teichmann and Babu[34], but also the result of the divergent effects of the TF interactions in activating or repressing the transcription of duplicated genes. For instance, regulatory systems where the TF is maintained but a different regulatory role is gained (either activation or repression) in one of the duplicated genes, such as the *E. coli gntK* and *idnK* gluconate kinase genes, which are involved in 6-phosphogluconate synthesis in the Entner–Doudoroff and pentose phosphate pathways, respectively. Although the same TFs, CRP, GntR, and IdnR, regulate all these genes, IdnR represses the transcription of gntK, whereas it activates the transcription of idnK[10]. This regulatory diversification allows plasticity of the TRN without the need to increase the number of interactions within it, whether this occurs only by varying the type of regulation (positive or negative) exerted by the TFs on their targets. Thus, it is possible that modulation will be one of the first steps towards evolutionary innovation at a biochemical level, perhaps as a step towards the modification of the entire metabolic pathway.

Conclusion and future prospects

The adaptation of bacteria to different ecological niches is a reflection of the reconfigurations that occur at the level of gene regulation. Evolutionary changes in the TRN have played an important role for contemporaneous organisms, and the increasing amount of genomic data allows us to delve more deeply into this topic. In this regard, the repertoire of TFs expands or contracts in a lineage-specific manner to adapt to the environmental needs of organisms, as has been observed in previous comparative studies. Therefore, the number and type of TFs present in the actual genomes reflect the cell responses to changing environments that have been encountered during their evolution. An example of how the environment determines the number of TFs in genomes is found in the marine cyanobacterium *Thrichodesmium erythraeum* and in the soil proteobacterium *Sinorhizobium meliloti*, two organisms that exhibit similar genome sizes, 7.7 and 6.7 Mb, respectively, but have a contrasting number of regulators, 69 in *T. erythraeum* and 390 in *S. meliloti*. This difference in the number of regulators may be caused by lifestyle, which is more demanding for *S. meliloti*, an organism that inhabits a variety of microenvironments, such as soil, rhizospheres, and plant root interiors, where there is a constant fluctuation of physical and chemical parameters, while *T. erythraeum* inhabits a more constant environment[15]. In addition, it has been observed that the nature of the host (animal or plant) influences the genome size and the nature of a microbe's TFs. In this regard, Santos *et al.*[35] found that animal-associated bacteria are mainly related to the LacI, DeoR, and Xre families, which are globally associated with carbon metabolism, while plant-associated bacteria show a tendency towards the Fur, Crp, and LuxR families, which are associated with ferric uptake, cell-cell signaling, and global regulation processes.

Thus, it is probable that the ancestral genetic networks we observe today were probably a small group of DBDs that, while conserving their structure, diverged into a large variety of TFs. More recently, TFs underwent many cycles of domain rearrangements, where dimerisation and ligand-binding domains were gained and lost at different times. Furthermore, they evolved across a series of single-gene duplications, thus generating networks of regulatory genes that are arranged into these modules. These events may be quite recent and lineage specific, as we have learned from the uneven distributions of some TF families[25]. In conclusion, diverse evolutionary forces have generated and modelled the TRNs, such as gene duplication, gene loss, changes in regulatory mechanisms (regulatory role modulation), acquisition of new activities, modular rearrangements, and finally, functional divergence. We believe that with the availability of more information, we will be able to understand in a more comprehensive

fashion the evolutionary dynamics associated with regulatory networks. In this context, new experimental approaches that combine techniques such as chromatin immunoprecipitation, microarray analysis, or next-generation sequencing are allowing scientists to explore the processes of transcriptional regulation in vivo and to discover interactions not previously described, thus providing new opportunities to identify new regulatory mechanisms beyond TFs or more complex global signaling networks, such as the virulence traits of *Pseudomonas aeruginosa*[36].

Acknowledgments

We thank Alejandro Huerta-Saquero, Silvia Tenorio-Salgado and the anonymous reviewer for their critical reading of the manuscript. Support from DGAPA-UNAM (IN-209511) is gratefully acknowledged. Work reported here was completed during a sabbatical leave of absence supported by CONACYT Fellowship (165772) to E.P-R. Support from DGAPA-UNAM Postdoctoral Fellowship is acknowledged to M.A.M-N.

References

1. Browning, D.F. and Busby, S.J. (2004) *Nat.Rev. Microbiol.,* **2**, 57-65.
2. Madan Babu, M. and Teichmann, S.A. (2003) *Nucleic Acids Res.,* **31**, 1234-1244.
3. Moreno-Campuzano, S., Janga, S.C. and Perez-Rueda, E. (2006) *BMC Genomics*, **7**, 147.
4. Brune, I., Brinkrolf, K., Kalinowski, J., Puhler, A. and Tauch, A. (2005) *BMC Genomics*, **6**, 86.
5. Tenorio-Salgado, S., Huerta-Saquero, A. and Perez-Rueda, E. (2011) *Comput. Biol. Chem.,* **35**, 341-346.
6. Perez-Rueda, E. and Janga, S.C. (2010) *Mol. Biol. Evol.,* **27**, 1449-1459.
7. Charoensawan, V., Wilson, D. and Teichmann, S.A. (2010) *Nucleic Acids Res.,* **38**, 7364-7377.
8. Perez-Rueda, E., Collado-Vides, J. and Segovia, L. (2004) *Comput. Biol. Chem.,* **28**, 341-350.
9. Bhardwaj, N., Carson, M.B., Abyzov, A., Yan, K.K., Lu, H. and Gerstein, M.B. (2010) *PLoS Comput. Biol.*, **6**, e1000755.
10. Keseler, I.M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Munis-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A.G., Mackie, A., Paulsen, I., Gunsalus, R.P. and Karp, P.D. (2011) *Nucleic Acids Res.,* **39**, D583-590.
11. Lee, J.S., Son, B., Viswanathan, P., Luethy, P.M. and Kroos, L. (2011) *J. Bacteriol.,* **193**, 1681-1689.
12. Mittal, S. and Kroos, L. (2009) *J. Bacteriol.,* **191**, 2753-2763.
13. Sierro, N., Makita, Y., de Hoon, M. and Nakai, K. (2008) *Nucleic Acids Res.,* **36**, D93-96.
14. Bell, S.D. (2005) *Trends Microbiol.,* **13**, 262-265.
15. Ulrich, L.E., Koonin, E.V. and Zhulin, I.B. (2005) *Trends Microbiol.,* **13**, 52-56.
16. Aravind, L. and Koonin, E.V. (1999) *Nucleic Acids Res.,* **27**, 4658-4670.

17. Chaikam, V. and Karlson, D.T. (2010) *BMB Rep.,* **43**, 1-8.
18. Itzkovitz, S., Tlusty, T. and Alon, U. (2006) *BMC Genomics,* **7**, 239.
19. Gotfredsen, M. and Gerdes, K. (1998) *Mol. Microbiol.,***29**, 1065-1076.
20. Ranea, J.A., Buchan, D.W., Thornton, J.M. and Orengo, C.A. (2004) *J. Mol. Biol.,* **336**, 871-887.
21. Ranea, J.A., Grant, A., Thornton, J.M. and Orengo, C.A. (2005) *Trends Genet.,* **21**, 21-25.
22. Cases, I., de Lorenzo, V. and Ouzounis, C.A. (2003) *Trends Microbiol.,* **11**, 248-253.
23. Cherry, J.L. (2003) *J. Theor. Biol.,* **221**, 401-410.
24. Janga, S.C. and Perez-Rueda, E. (2009) *Comput. Biol. Chem.,* **33**, 261-268.
25. Perez-Rueda, E., Janga, S.C. and Martinez-Antonio, A. (2009) *Mol. Biosyst.,* **5**, 1494-1501.
26. Miller, M.B. and Bassler, B.L. (2001) *Annu. Rev. Microbiol.,* **55**, 165-199.
27. Rigali, S., Schlicht, M., Hoskisson, P., Nothaft, H., Merzbacher, M., Joris, B. and Titgemeyer, F. (2004) *Nucleic Acids Res.,* **32**, 3418-3426.
28. Rivera-Gomez, N., Segovia, L. and Perez-Rueda, E. (2011) *Microbiology,* **157**, 2308-2318.
29. Wang, L., Wang, F.F. and Qian, W. (2011) *J. Genet. Genomics,* **38**, 279-288.
30. Shen-Orr, S.S., Milo, R., Mangan, S. and Alon, U. (2002) *Nat. Genet.,* **31**, 64-68.
31. Perez, J.C. and Groisman, E.A. (2009) *Cell,* **138**, 233-244.
32. Perez, J.C. and Groisman, E.A. (2009) *Proc. Natl. Acad. Sci. USA,* **106**, 4319-4324.
33. Martinez-Nunez, M.A., Perez-Rueda, E., Gutierrez-Rios, R.M. and Merino, E. (2010) *Microbiology,* **156**, 14-22.
34. Teichmann, S.A. and Babu, M.M. (2004) *Nat. Genet.,* **36**, 492-496.
35. Santos, C.L., Tavares, F., Thioulouse, J. and Normand, P. (2009) *FEMS Microbiol. Rev.,* **33**, 411-429.
36. Goodman, A.L. and Lory, S. (2004) *Curr. Opin. Microbiol.,* **7**, 39-44.