

Völf J-N (ed): Gene and Protein Evolution.
Genome Dyn. Basel, Karger, 2007, vol 3, pp 66–80

.....

General Trends in the Evolution of Prokaryotic Transcriptional Regulatory Networks

M. Madan Babu^{a,b}, S. Balaji^a, L. Aravind^a

^aNational Center for Biotechnology Information, National Institutes of Health, Bethesda, Md., USA; ^bMRC Laboratory of Molecular Biology, Hills Road, Cambridge, UK

Abstract

Gene expression in organisms is controlled by regulatory proteins termed transcription factors, which recognize and bind to specific nucleotide sequences. Over the years, considerable information has accumulated on the regulatory interactions between transcription factors and their target genes in various model prokaryotes, such as *Escherichia coli* and *Bacillus subtilis*. This has allowed the representation of this information in the form of a directed graph, which is commonly referred to as the transcriptional regulatory network. The network representation provides us with an excellent conceptual framework to understand the structure of the transcriptional regulation, both at local and global levels of organization. Several studies suggest that the transcriptional network inferred from model organisms may be approximated by a scale-free topology, which in turn implies the presence of a relatively small group of highly connected regulators (hubs or global regulators). While the graph theoretical principles have been applied to infer various properties of such networks, there have been few studies that have actually investigated the evolution of the transcriptional regulatory networks across diverse organisms. Using recently developed computational methods that exploit various evolutionary principles, we have attempted to reconstruct and compare these networks across a wide-range of prokaryotes. This has provided several insights on the modification and diversification of network structures of various organisms in course of evolution. Firstly, we observed that target genes show a much higher level of conservation than their transcriptional regulators. This in turn suggested that the same set of functions could be differently controlled across diverse organisms, contributing significantly to their adaptive radiations. In particular, at the local level of network structure, organism-specific optimization of the transcription network has evolved primarily via tinkering of individual regulatory interactions rather than whole scale reuse or deletion of network motifs (local structure). In turn, as phylogenetic diversification proceeds, this process appears to have favored repeated convergence to scale-free-like structures, albeit with different regulatory hubs.

Copyright © 2007 S. Karger AG, Basel

The pioneering studies by Jacob and Monod suggested the existence of regulatory proteins which bind to DNA elements upstream of other genes and control their expression. These regulatory proteins, termed transcription factors, respond to different signals and in turn activate or repress the expression of their target genes at appropriate instances [1–5]. Following these findings, several studies over the many years have accumulated a wealth of information on individual regulatory interactions mediated by these transcription factors in various model organisms [6–13]. More recently, there has been considerable interest and effort to assemble this information to derive what is termed the transcriptional regulatory network of an organism [14–16]. The topology of the transcriptional regulatory network is best modeled as a network or a graph with nodes representing transcription factors and target genes, and directed edges connecting the former to the latter [15–17].

Several recent studies on transcriptional networks of prokaryotes and eukaryotes have shown that the structure of such networks shows three levels of organization [15]. At the most basic level, the network contains individual regulatory interactions between transcription factors and targets (fig. 1). At the intermediate level, studies have shown multiple basic units to be organized into functionally distinct units called network motifs (fig. 1). Different types of motifs are defined based on patterns of interconnections between the basic units, and multiple copies of individual motif-types are found in different contexts within the network. Finally, at the highest level of organization, the set of all transcriptional regulatory interactions in a cell form the global structure and has been shown to have a hierarchical or a scale-free topology. In other words, such a global structure is characterized by the presence of a majority of transcription factors which regulate few genes and the presence of a few transcription factors, called the regulatory hubs, which regulate many genes (fig. 1).

While there has been significant progress, due to several experimental studies performed over many years, in unraveling the transcriptional regulatory networks of various model organisms such as *E. coli* and *B. subtilis*, much less information is available on the transcriptional networks of other prokaryotes. In order to gain a better understanding of the transcriptional regulatory network in other organisms, computational methods to extrapolate this information from model organisms to poorly studied organisms by exploiting the wealth of information in the form of publicly available completely sequenced genomes have been developed [18–29]. Such methods can be broadly grouped into two classes: (i) Orthology based methods: This approach exploits the basic principle that orthologous transcription factors regulate orthologous target genes in the different genomes. This method requires the transcriptional regulatory network for a reference organism and uses the protein sequences of transcription factors and target genes in the reference network to identify orthologs in the query

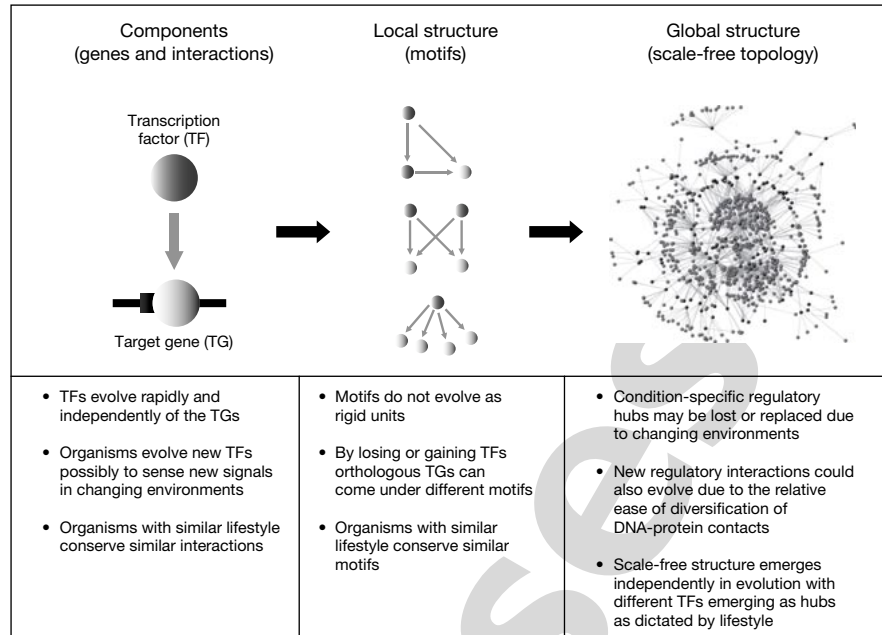


Fig. 1. Structure and evolution of transcriptional regulatory networks in prokaryotes. There are three levels of organization of network structure. (i) The basic unit is made up of a transcription factor, its target gene and a regulatory interaction represented as a directed arrow. (ii) At the local level, the basic unit forms network motifs, which are small patterns of interconnections with specific information processing ability. (iii) The set of all regulatory interactions in an organism, which is a representation of the transcriptional program of a cell, is referred to as the global structure of the network. The observed general evolutionary trends at the three levels of organizations of the network suggests that transcriptional regulatory networks in prokaryotes are very flexible and rapidly adapt to changes in environment by tinkering individual interactions to arrive at an organism specific optimal design.

organism in order to infer regulatory interactions in the genome of interest. (ii) Binding site profile based methods: This approach requires reliable information on the DNA binding site for a transcription factor. Further, it exploits the fact that presence of the same binding site upstream of different genes in a closely related species would imply a regulatory influence of an orthologous transcription factor on the expression of the nearby gene through the same binding site. Both these methods have their advantages and disadvantages; the former method allows prediction of conserved interactions and loss of interactions in distantly related organisms but does not facilitate discovery of novel targets for a given transcription factor. In contrast, the latter method allows detection of

novel targets for a transcription factor but is not applicable to distantly related genomes because DNA regulatory elements are shorter and evolve much faster than the protein-coding sequences, hence making detection of new interactions unreliable.

The availability of complete genome sequences of over 300 prokaryotes and the understanding of the structure of transcriptional regulatory networks have allowed us to address several fundamental questions on the origins and evolution of the transcriptional regulatory networks. In addition, the availability of the previously discussed methods has provided us with an opportunity to identify the distinct evolutionary trends in shaping of transcriptional regulatory networks at various levels of organization. In this chapter, we review the results from recent studies which have addressed these questions at various levels of resolution and present a summary of the general trends that can be discerned.

Evolution of Transcription Factors and Target Genes

Regulatory interactions between transcription factors and target genes could potentially evolve through two distinct modes: (i) in which both the transcription factor and the target gene co-evolve, i.e. present or absent as a pair or (ii) the transcription factor and the target gene evolve independently of each other. Our analysis of the conservation patterns (employing the orthology based method) of the genes and the regulatory interactions across 175 different prokaryotes revealed several interesting trends [29].

Using the *E. coli* transcriptional regulatory network, which consisted of 112 transcription factors, 755 target genes and 1,292 regulatory interactions, as the reference network, we found that the evolutionary retention of transcription factors in other organisms is lower than their target genes. The relatively low retention of transcription factors in other organisms does not mean that the regulatory influences on the more highly retained target genes are absent. We found that each organism has evolved its own set of transcriptional regulators that are not orthologous to other proteins suggesting innovation of regulatory proteins, possibly to sense new signals in changing environments and provide a new set of regulatory influences on target genes. Thus evolutionary forces appear to independently retain or discard transcription factors and their targets, with a higher frequency of loss or replacement of the former.

Several studies along these lines have recently demonstrated that there is a non-linear increase in the number of transcription factors encoded as the genome size of the organism increases [30–32]. These observations suggest that as genome size increases, more transcription factors are needed to regulate specialized groups of genes individually. Alternatively, it may also suggest the need

to integrate distinct inputs in order to introduce more layers in the regulatory hierarchy of metabolically or organizationally complex organisms with large genomes. These studies also revealed that (i) in parasites with small genomes, transcription factors have been lost due to absence of selective pressure for regulating target genes. These parasites could depend on the host cellular machinery for the fulfillment of roles performed by some of their proteins and (ii) in larger genomes, target genes are often controlled by additional regulators or regulators that are non-orthologous, so that there is an integration of a variety of different inputs that are typically dependent on the environmental niche of the organism [29, 32, 33]. For instance, in a number of phylogenetically distant free-living bacteria, including proteobacteria, *Bacillus subtilis* and *Streptomyces*, there is an expansion of the so-called one-component transcription factors of the LysR family, which sense a wide range of small molecule ligands. However, there are no homologs of such transcription factors in any of their close relatives, which are obligate pathogens. This is consistent with the need to sense a similar set of environmental metabolites by all of the above-mentioned free-living bacteria.

At the level of regulatory interactions, we found that organisms which are phylogenetically distantly related but share similar environmental life-style tend to significantly conserve regulatory interactions, hinting a prominent influence for environment or life-style in selecting for their conservation. For example, bacteria with comparable genome size, such as several species of *Bacillus*, *Corynebacterium* and *Mycobacterium*, whose principal habitat is the soil, conserve orthologous regulatory interactions. Likewise, the obligate or intracellular parasites from diverse bacterial clades, namely *Mycoplasma*, *rickettsiae* and *chlamydiae*, conserve similar regulatory interactions. To test the generality of this observation, an index which measures similarity in network structure and lifestyle (LSI) was developed [29] which revealed the existence of a strong evolutionary trend: organisms belonging to the same lifestyle have a significantly higher number of regulatory interactions in common in comparison to organisms from other lifestyle classes.

In addition, analysis of the conserved regulatory interactions in the different genomes allowed us to speculate about the components of the ancestral networks in the different phylogenetic lineages. The common ancestor of archaea and eubacteria appears to have had quite a few global regulatory proteins (with more than 14 target genes) such as Crp (cAMP receptor protein/regulator), Fnr (fumarate nitrate reduction regulator) and Lrp (regulator for leucine regulon). The predicted ancestral network at this level might have contained up to 62 TFs that regulate genes required for basic processes that sustain life, which include regulators for genes involved in purine biosynthesis, fructose utilization, xylose utilization, iron uptake, fatty acid biosynthesis, anaerobic respiration and

amino acid biosynthesis. There has been gain of specific regulatory systems in course of eubacterial evolution. These include transcriptional regulators that can sense a variety of different sugar molecules and their target genes that utilize these sugars (e.g. mellibiose, mannitol, glucitol, galactose, etc.) to generate energy. Within particular bacterial lineages, e.g. the firmicute lineage, which includes the endospore-forming bacteria and actinobacteria, appears to have lost regulatory systems that can utilize L-idonate, and the sulphur utilization system. Further, in the proteobacteria and cyanobacteria, there have been multiple instances where various regulatory systems have been lost in the different lineages. These results point to the possibility that several regulators were present in the ancestral genome but have been lost, displaced or retained as organisms colonized and adapted to new environmental niches.

Evolution of the Local Network Structure

Regulatory networks can be fragmented into fundamental regulatory sub-systems or motifs which when put together reconstruct the entire network. In the case of the *E. coli* network, three types of motifs (1) feed-forward motif, (2) single input motif and (3) multiple-input motifs have been discerned through a combination of computational and a series of experimental studies aimed at understanding their functions [34–39]. Such studies have elucidated that the feed-forward motif could ensure regulation of target genes only when a persistent signal is received, thereby filtering noise or fluctuation in the input signal. It was also demonstrated that the single input motif could co-ordinate global changes in gene expression and could enforce an order in the patterns of gene expression of its targets and that the multiple input motif could integrate different signals and hence could differentially regulate the target genes.

In principle evolution of these network motifs could follow any of the following trajectories: (i) a trend where all interactions in a network motif are conserved in other organisms or (ii) a trend where motifs are not conserved as complete units due to which individual interactions may be lost or gained during the course of evolution. Given that the network motifs have specific information processing ability, one might expect these motifs to be conserved as relatively rigid units (all the components of the motifs are conserved) once they have emerged. However, analysis of the conservation patterns of these network motifs across the 175 genomes [29] revealed the contrary that network motifs are not conserved as complete units in other organisms. At a first glance, it was surprising to find that organisms which were evolutionary close did not conserve regulatory network motifs whereas several organisms which were distantly related conserved orthologous network motifs (fig. 2a). For example, Fnr

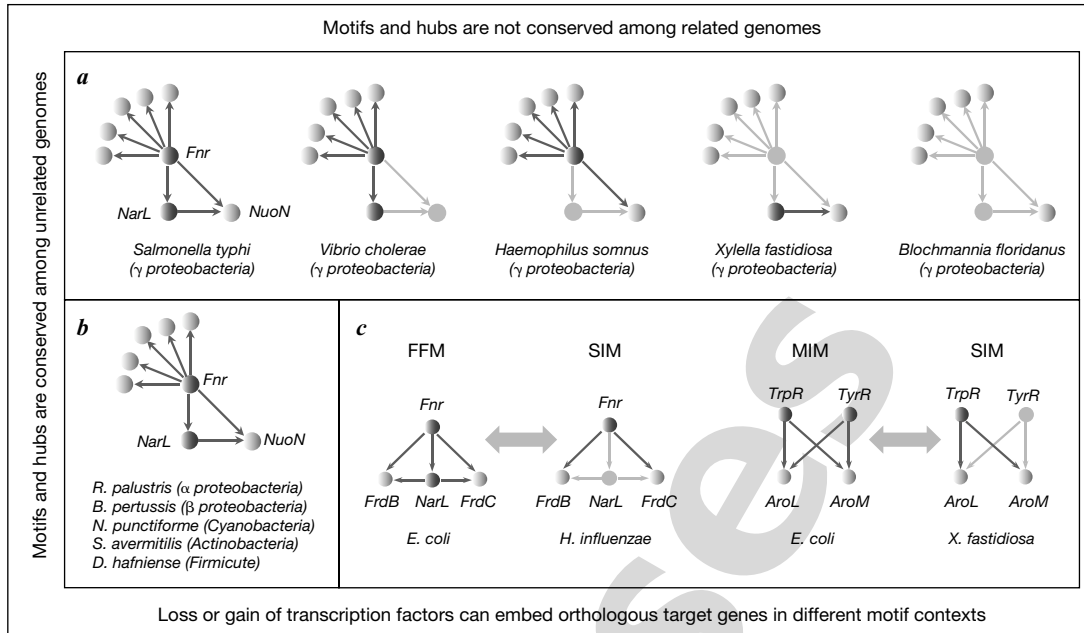


Fig. 2. Evolution of network motifs. **a** A feed-forward motif formed by transcription factors Fnr, NarL and Nuon in *E. coli* is completely conserved in a closely related genome, *Salmonella typhi*, but not in other gamma-proteobacterial genomes. **b** Distantly related organisms that have preserved all interactions in the regulatory motif and that have conserved the regulatory hub, Fnr. **c** Analysis of partially conserved motifs revealed that by losing (or gaining) specific transcription factors, orthologous genes in different genomes could be embedded in different motif contexts. Thus evolution tinkers with specific regulatory interactions when orthologous genes in organisms living in a different environment need to be expressed differently. In these figures the TFs and TGs are represented by dark grey circles and light grey circles, respectively, while white circles denote their absence.

(a global regulator, activated during low oxygen levels), NarL (transcriptional regulator of a two-component signal transduction system) and Nuon (subunit of the NADH dehydrogenase complex I) form a feed-forward motif in *E. coli*, which is not completely conserved as a unit in other gamma proteobacterial genomes. In contrast, all the interactions in this motif are conserved in several distantly related genomes such as the beta-proteobacterium *B. pertussis* and the firmicute *D. hafniense* (fig. 2b). Further careful analysis revealed the role of the environment in shaping the structure of these network motifs. It was found that in instances where the network motifs were not conserved in closely related organisms, they had significant differences in their life-styles. Strikingly, in

instances where distantly related organisms were found to conserve regulatory network motifs, it was seen that they had a considerable similarity in their lifestyle. A comprehensive analysis to assess the generality of this observation where organisms were grouped according to their lifestyle similarity and assessed for similarity in their network motif content revealed a statistically significant trend that organisms with similar lifestyle tend to regulate their target genes by means of similar network motifs [27].

It was not immediately clear how local network structure or network motifs of organisms evolve with diversification of life-style or environments. A case by case analysis of some of the partially conserved regulatory network motifs of organisms living in different environments revealed that by losing or gaining individual transcription factors, orthologous target genes could be potentially expressed in different ways according to the requirements of the organism. This meant that by losing or gaining individual regulatory proteins, organisms living in different environments can regulate orthologous target genes through different network motifs [29]. For instance, genes which are regulated through a feed-forward motif (FFM) can be regulated as a part of a single input motif (SIM) by losing a transcription factor (fig. 2c). Note that the regulation of a gene through a FFM would ensure that the target gene expression is not sensitive to fluctuations in input signals. Whereas regulation through a SIM would ensure expression of target genes as long as there is some input signal, for instance, the presence of a particular metabolite. Likewise, target genes regulated through a multiple input motif (MIM) in one organism can come to be regulated through a single input motif (SIM) by losing one of the transcription factors (fig. 2c).

For example, in *E. coli*, which is adapted to a lifestyle with largely fixed aerobic and anaerobic phases, the fumarate reductase genes (FrdB and FrdC, which converts fumarate to succinate under anaerobic conditions to derive energy) are not expressed unless there is a persistent signal for lack of oxygen received through a feed forward motif involving both Fnr and NarL. In contrast, *Haemophilus influenzae*, which encounters rapid redox fluctuations during host infection and needs to regulate the fumarate reductase genes more quickly than *E. coli*, appears to depend solely on Fnr for the response by employing a single input motif. Thus by losing a transcription factor, genes that are tightly regulated through a FFM can be regulated in a much simpler manner (fig. 2c).

These observations reveal an important principle in the evolution of network motifs that orthologous genes in related organisms living in different environments may acquire distinct patterns of gene expression by embedding them in appropriate motif context in order to adapt better to changing environments [40–44, 29]. Our findings also indicate that different organisms arrive at the best possible solutions to regulate the same gene by tinkering specific regulatory interactions in order to optimize expression levels rather than by duplicating

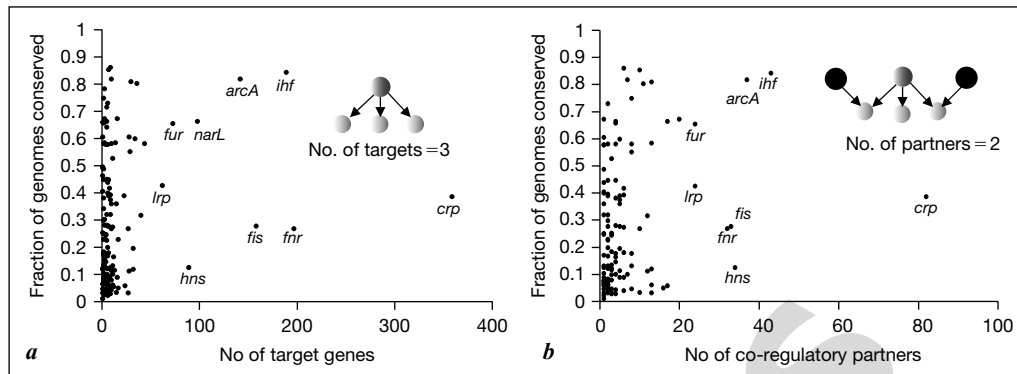


Fig. 3. Conservation v/s connectivity plot for the transcription factors in the transcriptional network of *E. coli*. **a** Fraction of the 330 genomes conserved against the number of target genes for a transcription factor. The connectivity in terms of the number of target genes for the transcription factor shown in black is three. **b** Fraction of the genomes conserved against the number of regulatory partners for a given transcription factor. The number of co-regulatory proteins for the transcription factor shown in black is two as it co-regulates its target genes with two other transcription factors. *arcA* is anaerobic respiration regulatory protein, *ihf* is integral host factor and *fur* is ferric uptake regulator.

groups of genes which are already a part of a motif. In this context, other studies on duplicated genes within the transcriptional regulatory network of *E. coli* and yeast [45–47] have shown that network motifs have not evolved by duplication of complete ancestral motifs lending support to our interpretation that the same interactions, which is a part of a motif in one organism, could have existed in different regulatory contexts in their ancestral genomes.

Evolution of the Global Network Structure

Results from graph theoretical studies and the fact that global regulatory hubs control the expression of several genes suggested that such hubs would assume importance in transcriptional networks and hence be more conserved in evolution than other transcription factors. In our investigation on the evolution of the global structure of prokaryotic transcriptional regulatory networks [29], we observed that transcriptional regulatory hubs are not preferentially retained than any other transcription factors in the network (fig. 3a). Careful analysis of the transcription factors which were lost in other organisms revealed that these are regulatory hubs that were largely condition specific (*narL*, *crp*, etc.) and

hence were lost in instances where the organism would not experience such a condition. For instance, an organism which has been adapted to live in an aerobic environment could dispense global regulators which are required only under anaerobic condition in order to optimize its genome content and to minimize resources spent by reducing expression of unwanted proteins. Our work also revealed that non-global regulatory transcription factors which control expression of specific regulatory systems are also largely dispensable if there is no selective advantage. For example, in the opportunistic pathogen *P. aeruginosa*, which actively utilizes phenolic compounds, the transcriptional regulators, MhpR, HcaR and FeaR, can sense the compounds and activate target genes that encode enzymes involved in their catabolism. However, more obligate pathogens like *Staphylococcus aureus* and *Campylobacter jejuni*, which do not typically face phenolic compounds in their natural niches, lack both the regulators and their target genes for the utilization of these aromatic compounds. Thus it appears that the absence of any selective pressure to maintain a regulatory protein would render even the global regulatory hubs dispensable, just like any other transcription factor in an organism.

In a recent study on the yeast transcriptional regulatory network [48, 49], we showed that global regulators can be of two distinct types: (i) those that regulate several of their targets in an autonomous manner and (ii) those that integrate signals through different transcription factors and hence combinatorially regulate the expression of their targets. In the light of our finding that there are two classes of global regulatory hubs, we assessed the existence of a trend that global regulators which tend to regulate several targets by themselves would be evolutionarily more conserved than those hubs which co-regulate with several other transcription factors. Our analysis showed that there is no such trend and that both classes of regulatory hubs tend to evolve like any other transcription factor in the genome (fig. 3b). Upon a closer look, we observed that in the *E. coli* transcriptional network, the regulatory hubs which had many target genes were also, in general, the ones which also have many co-regulatory partners indicating that autonomous hubs are far less in number than the integrator-type regulatory hubs.

Given that the global regulatory hubs are not conserved in evolution, we compared the experimentally characterized regulatory network of *E. coli* and *B. subtilis* to understand if there were any differences in the overall topology of the networks from organisms living in different environments. Even though our analysis revealed that the topology of both the networks adopted a similar scale-free structure, it also pointed to the fact that the proteins which emerge as global regulatory hubs in the two organisms are not evolutionarily related. For example, CcpA (which is activated by phosphorylation events) and Crp (which is activated by the presence of cAMP) are the two regulatory hubs in *B. subtilis*

and *E. coli* respectively controlling many genes involved in carbon metabolism. Both have very different modes of regulation and are not evolutionarily related. This suggests that regulatory hubs have been independently innovated to regulate orthologous target genes in organisms living in different environments. These observations provide strong support that the hierarchical structure of these networks has converged to a similar scale-free topology, albeit with independently recruited regulatory hubs. We believe that such an emergence of evolutionarily unrelated proteins to the status of a regulatory hub can be explained because the binding affinity and specificity of a transcription factor and its target site can be affected by relatively small changes in the DNA-binding interface of the transcription factor, or in the binding site [50]. As a result DNA-binding domains could evolve new target sites relatively easily, resulting in rapid de novo emergence of new transcriptional interactions.

Taken together our observations suggest general principles of evolution at the level of global regulatory proteins: (i) transcription factors which are condition specific, be it global regulatory hubs or transcriptional regulators of specific systems, are dispensable in the absence of any selective pressure to maintain them in an organism. (ii) The extent of advantage conferred by orthologous transcription factors to the fitness of an organism might vary across organisms depending on the environment and hence during the course of evolution different proteins may emerge as regulatory hubs in organisms colonizing different niches. (iii) Though different proteins emerge as regulatory hubs, transcriptional regulatory networks tend to approximate a scale-free topology, suggesting that this is a global property, which is enforced entirely independently of the evolutionary forces on the constituent elements of the network [29].

Conclusion

Transcriptional networks, which can be studied at three distinct levels of organization, have been shaped by disparate forces acting at different levels. At the level of the components which comprise these networks, i.e. transcription factors, target genes and regulatory interactions, we observe that (i) the transcription factors complement changes more rapidly than the target genes, with organisms colonizing different ecological niches by evolving their own set of novel transcriptional regulators. This suggests that a major factor in the emergence of new life-styles is the evolution of distinct repertoires of transcription factors, which probably integrate new input signals. (ii) Organisms with similar lifestyle tend to possess similar regulatory interactions.

In terms of trends which are seen in the evolution of network motifs, we note that (i) network motifs which have the ability to finely regulate the expression

of the target genes are not conserved as rigid units across the different organisms. However, organisms with similar lifestyle tend to regulate orthologous target genes through similar network motifs suggesting that regulation of genes through appropriate motifs could confer advantage to an organism. (ii) We also note that by losing or conserving specific transcriptional regulators, orthologous genes in different genomes can be incorporated within different regulatory contexts and can thereby easily exhibit different patterns of gene expression. This suggests that natural selection tinkers with individual interactions to arrive at an optimal design to regulate a gene in a given organism.

Finally, at the level of the global network structure, we note that (i) conservation of transcription factors is independent of the number of target genes they regulate or the number of other transcription factors with which a given regulator interacts. Instead, it appears that the determining factor for the retention of a transcriptional regulator appears to be the lifestyle of the organism. (ii) Additionally, it appears that the same transcription factor can have differential functional relevance for organisms living in different environments due to which evolutionarily unrelated proteins could emerge as hubs in different organisms during the course of evolution. Though different proteins emerge as regulatory hubs, the overall scale-free topology is maintained suggesting that such a structure has evolved convergently and is an emergent property in evolution.

The computational methods developed in our study and those of others, when integrated together with the results from experimental studies which employ recently developed techniques (such as DamID [51], ChIP-chip [52–54], CLIP [55], 1- and 3-hybrid experiments [56] and 2D-EMSA [57]) could complement each other in uncovering the details of transcriptional control in poorly characterized organisms. The predictions from such integrative approaches might allow better design of experiments for biochemical engineering and anti-pathogen therapeutics.

Supplementary URL

Supplementary information detailing our orthology based approach and the predictions of transcriptional networks and transcription factors are available at <http://www.mrc-lmb.cam.ac.uk/genomes/madanm/evdy/>

Acknowledgements

The authors gratefully acknowledge the Intramural research program of National Institutes of Health, USA for funding their research.

References

- 1 Jacob F, Monod J: Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 1961;3: 318–356.
- 2 Ptashne M, Jeffrey A, Johnson AD, Maurer R, Meyer BJ, et al: How the lambda repressor and cro work. *Cell* 1980;19:1–11.
- 3 Takeda Y, Ohlendorf DH, Anderson WF, Matthews BW: DNA-binding proteins. *Science* 1983;221:1020–1026.
- 4 Pabo CO, Sauer RT: Protein-DNA recognition. *Annu Rev Biochem* 1984;53:293–321.
- 5 Browning DF, Busby SJ: The regulation of bacterial transcription initiation. *Nat Rev Microbiol* 2004;2:57–65.
- 6 Svetlov VV, Cooper TG: Review: compilation and characteristics of dedicated transcription factors in *Saccharomyces cerevisiae*. *Yeast* 1995;11:1439–1484.
- 7 Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, et al: AGRIS: *Arabidopsis* gene regulatory information server, an information resource of *Arabidopsis* cis-regulatory elements and transcription factors. *BMC Bioinformatics* 2003;4:25.
- 8 Makita Y, Nakao M, Ogasawara N, Nakai K: DBTBS: database of transcriptional regulation in *Bacillus subtilis* and its contribution to comparative genomics. *Nucleic Acids Res* 2004; 32(Database issue):D75–D77.
- 9 Martinez-Bueno M, Molina-Henares AJ, Pareja E, Ramos JL, Tobes R: BacTregulators: a database of transcriptional regulators in bacteria and archaea. *Bioinformatics* 2004;20:2787–2791.
- 10 Gonzalez AD, Espinosa V, Vasconcelos AT, Perez-Rueda E, Collado-Vides J: TRACTOR_DB: a database of regulatory networks in gamma-proteobacterial genomes. *Nucleic Acids Res* 2005;33 (Database issue):D98–D102.
- 11 Baumbach J, Brinkrolf K, Czaja LF, Rahmann S, Tauch A: CoryneRegNet: an ontology-based data warehouse of corynebacterial transcription factors and regulatory networks. *BMC Genomics* 2006;7:24.
- 12 Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, et al: TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 2006;34 (Database issue):D108–D110.
- 13 Salgado H, Gama-Castro S, Peralta-Gil M, Diaz-Peredo E, Sanchez-Solano F, et al: RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res* 2006;34(Database issue):D394–D397.
- 14 Thieffry D, Huerta AM, Perez-Rueda E, Collado-Vides J: From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays* 1998;20:433–440.
- 15 Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA: Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* 2004;14:283–291.
- 16 Barabasi AL, Oltvai ZN: Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;5:101–113.
- 17 Albert R: Scale-free networks in cell biology. *J Cell Sci* 2005;118:4947–4957.
- 18 McGuire AM, Hughes JD, Church GM: Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Res* 2000;10:744–757.
- 19 Tan K, Moreno-Hagelsieb G, Collado-Vides J, Stormo GD: A comparative genomics approach to prediction of new members of regulons. *Genome Res* 2001;11:566–584.
- 20 Rajewsky N, Socci ND, Zapotocky M, Siggia ED: The evolution of DNA regulatory regions for proteo-gamma bacteria by interspecies comparisons. *Genome Res* 2002;12:298–308.
- 21 Alkema WB, Lenhard B, Wasserman WW: Regulog analysis: detection of conserved regulatory networks across bacteria: application to *Staphylococcus aureus*. *Genome Res* 2004;14: 1362–1373.
- 22 Gao F, Foat BC, Bussemaker HJ: Defining transcriptional networks through integrative modeling of mRNA expression and transcription factor binding data. *BMC Bioinformatics* 2004;5:31.
- 23 Rodionov DA, Dubchak I, Arkin A, Alm E, Gelfand MS: Reconstruction of regulatory and metabolic pathways in metal-reducing delta-proteobacteria. *Genome Biol* 2004;5:R90.

- 24 Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, et al: Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. *Genome Res* 2004;14:1107–1118.
- 25 Espinosa V, Gonzalez AD, Vasconcelos AT, Huerta AM, Collado-Vides J: Comparative studies of transcriptional regulation mechanisms in a group of eight gamma-proteobacterial genomes. *J Mol Biol* 2005;354:184–199.
- 26 Rodionov DA, Dubchak IL, Arkin AP, Alm EJ, Gelfand MS: Dissimilatory metabolism of nitrogen oxides in bacteria: comparative reconstruction of transcriptional networks. *PLoS Comput Biol* 2005;1:e55.
- 27 Barrett CL, Palsson BO: Iterative reconstruction of transcriptional regulatory networks: an algorithmic approach. *PLoS Comput Biol* 2006;2:e52.
- 28 Lozada-Chavez I, Janga SC, Collado-Vides J: Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res* 2006;34:3434–3445.
- 29 Madan Babu M, Teichmann SA, Aravind L: Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol* 2006;358:614–633.
- 30 van Nimwegen E: Scaling laws in the functional content of genomes. *Trends Genet* 2003;19:479–484.
- 31 Ranea JA, Buchan DW, Thornton JM, Orengo CA: Evolution of protein superfamilies and bacterial genome size. *J Mol Biol* 2004;336:871–887.
- 32 Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM: The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* 2005;29:231–262.
- 33 Martinez-Antonio A, Janga SC, Salgado H, Collado-Vides J: Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*. *Trends Microbiol* 2006;14:22–27.
- 34 Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, et al: Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* 2001;292:2080–2083.
- 35 Mangan S, Alon U: Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* 2003;100:11980–11985.
- 36 Dekel E, Mangan S, Alon U: Environmental selection of the feed-forward loop circuit in gene-regulation networks. *Phys Biol* 2005;2:81–88.
- 37 Kalir S, Mangan S, Alon U: A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. *Mol Syst Biol* 2005;1:2005.0006.
- 38 Mangan S, Itzkovitz S, Zaslaver A, Alon U: The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *J Mol Biol* 2006;356:1073–1081.
- 39 Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U: Plasticity of the cis-regulatory input function of a gene. *PLoS Biol* 2006;4:e45.
- 40 Elena SF, Lenski RE: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* 2003;4:457–469.
- 41 Bjornstad ON, Harvill ET: Evolution and emergence of Bordetella in humans. *Trends Microbiol* 2005;13:355–359.
- 42 Dekel E, Alon U: Optimality and evolutionary tuning of the expression level of a protein. *Nature* 2005;436:588–592.
- 43 Fong SS, Joyce AR, Palsson BO: Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Res* 2005;15:1365–1372.
- 44 Babu MM, Aravind L: Adaptive evolution by optimizing expression levels in different environments. *Trends Microbiol* 2006;14:11–14.
- 45 Conant GC, Wagner A: Convergent evolution of gene circuits. *Nat Genet* 2003;34:264–266.
- 46 Madan Babu M, Teichmann SA: Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res* 2003;31:1234–1244.
- 47 Teichmann SA, Babu MM: Gene regulatory network growth by duplication. *Nat Genet* 2004;36:492–496.
- 48 Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L: Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J Mol Biol* 2006;360:213–227.
- 49 Balaji S, Iyer LM, Aravind L, Babu MM: Uncovering a hidden distributed architecture behind scale-free transcriptional regulatory networks. *J Mol Biol* 2006;360:204–212.

- 50 Luscombe NM, Thornton JM: Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. *J Mol Biol* 2002;320:991–1009.
- 51 van Steensel B, Henikoff S: Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* 2000;18:424–428.
- 52 Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO: Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 2001;409:533–538.
- 53 Horak CE, Snyder M: ChIP-chip: a genomic approach for identifying transcription factor binding sites. *Methods Enzymol* 2002;350:469–483.
- 54 Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, et al: Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 2002;298:799–804.
- 55 Ule J, Jensen K, Mele A, Darnell RB: CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* 2005;37:376–386.
- 56 Drees BL: Progress and variations in two-hybrid and three-hybrid technologies. *Curr Opin Chem Biol* 1999;3:64–70.
- 57 Woo AJ, Dods JS, Susanto E, Ulgiati D, Abraham LJ: A proteomics approach for the identification of DNA binding activities observed in the electrophoretic mobility shift assay. *Mol Cell Proteomics* 2002;1:472–478.

M. Madan Babu and L. Aravind
National Center for Biotechnology Information
National Institutes of Health, Bethesda, MD 20894 (USA)
Tel. +44 1223 402041, Fax +44 1223 213556
E-Mail madanm@mrc-lmb.cam.ac.uk or aravind@ncbi.nlm.nih.gov