The two faces of short-range evolutionary dynamics of regulatory modes in bacterial transcriptional regulatory networks

S. Balaji* and L. Aravind

Summary

Studies on the conservation of the inferred transcriptional regulatory network of prokaryotes have suggested that specific transcription factors are less-widely conserved in comparison to their target genes. This observation implied that, at large evolutionary distances, the turnover of specific transcription factors through loss and non-orthologous displacement might be a major factor in the adaptive radiation of prokaryotes. However, the recent work of Hershberg and Margalit[1] suggests that, at shorter phylogenetic scales, the evolutionary dynamics of the bacterial transcriptional regulatory network might exhibit distinct patterns. The authors find previously unnoticed relationships between the regulatory mode (activation or repression), the number of regulatory interactions and their conservation patterns in γ-proteobacteria. These relationships might be shaped by the differences in the adaptive value and mode of operation of different regulatory interactions. ***BioEssays*** **29:625–629, 2007. © 2007 Wiley Periodicals, Inc.**

Introduction

Transcription regulation is mediated by specific transcription factors (TFs), which regulate a particular set of target genes (TGs), by specifically recognizing and binding their promoters. Regulation by specific TFs can either cause ''activation'' or ''repression'', which respectively corresponds to increase or decrease of mRNA expression levels with respect to the base line. Over the years, individual studies as well as high-throughput methods have generated an enormous wealth of information on the regulatory inputs provided by specific TFs to their target genes. This has allowed the assembly of transcriptional regulatory interactions and their modes on the genome scale for the prokaryotic model organism, *E.coli* K12. These data have been made publicly available as RegulonDB (URL: http://regulondb.ccg.unam.mx/index. html),[2] and is being widely used as a base for genomic studies on the structure and evolution of transcription regulation. Typically this information is represented as a network or ordered graph, termed the transcriptional regulatory network (TRN), with two kinds of nodes, namely the TFs and TGs.[3,4] Distribution of regulatory interactions of TFs has been shown to be approximated by a power-law decay.[5,6] This implies that the *E. coli* TRN has a scale-free topology with a few TFs (hubs) regulatory a large number of TGs, while the rest of the TFs have a limited number of TGs.

Some earlier studies on the evolution of the inferred TRN across a phylogenetically wide range of completely sequenced prokaryotic genomes have suggested that TFs and TGs are retained or lost independently of each other.[7,8] In general TGs were found to be maintained to a greater extent than their upstream TFs. It was also observed that hubs were not preferentially retained over TFs with a small number of TGs. These observations on bacterial TRNs suggested that they might be highly flexible, with a notable turnover in the course of evolution of the specific TFs via loss and non-orthologous displacement. This provided a possible model for the adaptive radiations of bacteria, wherein the target genes are maintained across lineages but their regulatory inputs are drastically altered by the turnover of TFs. However, the details of this process at close phylogenetic distances remained unclear. Hershberg and Margalit attempted to understand this by focusing on the γ-proteobacteria using a wealth of recently available genomic data for these organisms.[1] Interestingly, the authors found that, within enterobacterial lineage of γ-proteobacteria, co-evolution of TFs and their TGs is related to the mode of regulatory interactions between them, i.e. activation or repression. Specifically, repressors, unlike activators, tend to co-evolve tightly with their TGs. Repressors

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD.
*Correspondence to: S. Balaji, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894. E-mail: sbalaji@ncbi.nlm.nih.gov
DOI 10.1002/bies.20600
Published online in Wiley InterScience (www.interscience.wiley.com).

with many targets are preferentially conserved as compared to activators with many targets.

Categorization of TFs and gross evolutionary dynamics of TFs in γ-proteobacteria

For the purpose of their analysis, the authors categorized TFs using two basic schemes: (1) evolutionary conservation and (2) regulatory modes. In the first scheme, which is based on identification of orthologs (Fig. 1), TFs were classified as "widely present" in 30 γ-proteobacterial genomes, "Entero-present" referring to their wide presence only in the enterobacterial lineage (which includes *E.coli* and its closest relatives) and "Entero-absent", which refers to TFs that are absent in this lineage. In the second scheme, the authors used the information on individual regulatory modes of TFs provided by RegulonDB to categorize them as activators, repressors and dual regulators (which both activate or repress different sets of TGs). The *E.coli* TRN used in their study contained 143 TFs, 1048 TGs and 2285 regulatory interactions. Based on the number of regulatory interactions of TFs in the TRN, the authors found that nearly half of the TFs (72 out of 143) in the TRN have more than 5 TGs. These 72 TFs were then categorized as those with "many" targets and remaining 71 TFs were categorized as ones with "few" targets. The numbers of TFs respectively classified into "widely present", "Entero-present" and "Entero-absent" were 13, 65 and 65.

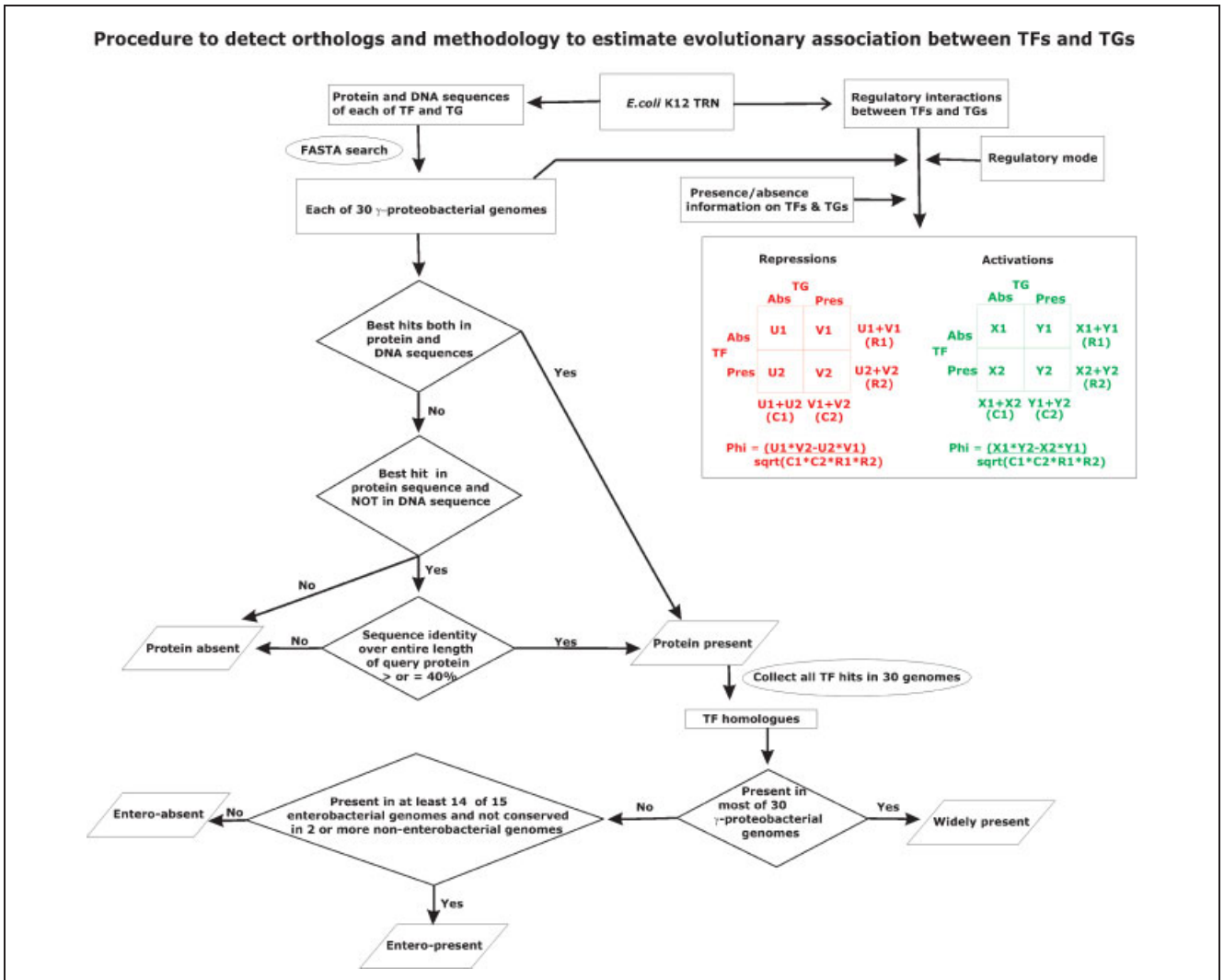

Figure 1. Procedure used by Hershberg and Margalit[1] in detecting orthologs. Also shown is methodology used by the authors for evaluating the extent of evolutionary associations between TFs and TGs in the *E. coli* TRN for both activating and repressive mode of regulatory interactions in each of the 30 γ-proteobacterial genomes. The four entries in both red and green tables, which respectively correspond to activations and repressions, contain a number of instances wherein both the TGs and their TFs were conserved, neither of them was conserved or only one of them was conserved.

The number of activators, repressors and dual regulators were 68, 49 and 26 respectively.

These observations and analysis of the different categories suggested several notable features regarding the evolution of the TRN in γ-proteobacteria.

(1) The low number of TFs in the "widely present" category suggests that only a relatively small proportion of TFs are widely conserved even within γ-proteobacteria lineage, consistent with the earlier studies on inferred TRNs of prokaryotes.
(2) TFs that solely activate or repress are much more prevalent than TFs that perform both the roles.
(3) Activators and repressors were not widely conserved compared to dual regulators (Fig. 2).
(4) Activators and repressors with few targets display comparable phyletic spread within the enterobacterial lineage (Fig. 2).
(5) Among TFs with many targets, the number of repressors conserved within the enterobacterial lineage is more than twice that of activators (Fig. 2).

Preferential retention of repressors with many targets has not been reported in earlier large-scale comparisons of TFs in bacteria. Hence, it appears that, in short-phylogenetic ranges, such as the enterobacterial clade, there might be selective constraints to conserve repressors as compared to activators.

Co-evolution of TFs and their target genes

The authors devised a procedure to specifically decipher the relationship, if any, between regulatory mode (activation, repression) and co-evolution of TFs and their TGs (Fig. 1). They omitted 95 regulatory interactions involving rRNA, tRNA and ncRNA target genes and concentrated only on protein-coding genes. The procedure employed could be divided in three steps.

(1) For each of the 30 γ-proteobacterial genomes, the information on presence or absence of TFs and TGs were recorded and mapped on to all the TF and TG regulatory interactions in the *E. coli* TRN.
(2) This information was combined with data on mode of regulatory interactions in the *E. coli* TRN, two 2×2 tables for each genome—one for activation and the other for repression. In each of the four cells of these tables, the following entries were recorded: the number of interactions with (i) both TF and its TG were conserved, (ii) both TF and its TG were not conserved, (iii) only TF was conserved and (iv) only TG was conserved (Fig. 1).
(3) Then a statistical measure, the phi-coefficient, was used to assess the extent of association or co-evolution between TFs and their targets in each of these tables. Analogous to Pearson correlation coefficient, phi-coefficient takes up values between -1 to $+1$. The higher the value of phi-coefficient the greater is the extent of co-occurrence

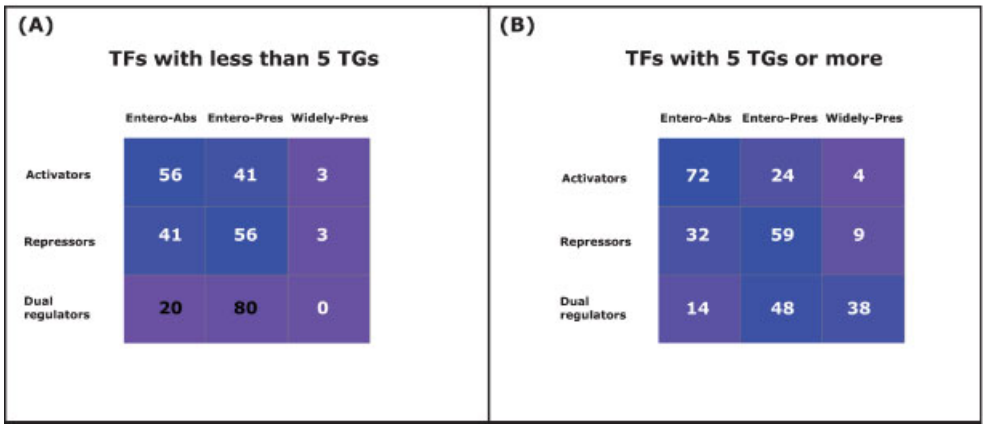

Figure 2. Matrix plots depicting percentage of activators, repressors and dual regulators in each category of evolutionary conservation (see Fig. 1). The percentage values have been taken from the original data. The plot on the left (A) corresponds to TFs with less than 5 TGs and (B) corresponds to TFs with 5 or more target genes. The values shown in black denote that there are only few representations in that category. The dark blue and violet boxes indicate high and low extent of representation of TFs. Entero-Abs, Entero-Pres and Widely-Pres denote respectively "Entero-absent", "Entero-present" and "widely present" category of TFs. The matrix plots were made using matrix2png program.[10]

between TFs and TGs. The statistical significance of each of this phi-coefficient measure was assessed using a *P*-value measure, which indicates the chance of getting the given value of phi-coefficient purely by chance. The *P*-values <0.05 were considered to reflect statistically significant phi-coefficients.

The authors found that the phi-coefficients for repressive interactions are significantly higher than those for activating interactions in all 15 enterobacterial genomes. Thus, it was inferred that there is higher propensity for repressors and their TGs to be present or absent simultaneously. Further, the authors also evaluated the probability of conservation of TGs given that their activators or repressors are not conserved. The probability measures suggested that, in most of enterobacterial genomes, repressors have significantly less tendency as compared to activators to be lost if their target genes are conserved. Thus, even though on larger evolutionary scales TFs in general are less-widely conserved in comparison to TGs,[7] on smaller scales, repressors tend to be co-maintained with their TGs.

Potential implications of TRN evolution in small phylogenetic ranges

The "zoomed-in" view of evolutionary dynamics of bacterial TRN, which is restricted to enterobacterial lineage, present certain differences vis-à-vis the global evolutionary picture. We suspect that there might be a potential relationship between the higher conservation of repressors with many targets and the tighter co-evolution of repressors and their TGs. While there might be multiple repressors acting at a given promoter region, it appears that they typically provide regulatory inputs "in series", responding to different physiological or signaling states. Thus, many of the repressors provide single inputs to their target genes, each at a given time (Fig. 3). However, activators tend to allow more-rapid uses of resources, and appear to relatively more often provide combinatorial regulatory inputs in parallel (Fig. 3). Thus, for a given condition, retention of a target gene and not its repressor could allow unnecessary transcriptional responses in the organism, and levy a cost which could potentially reduce its fitness.[9] Further, the more the number of TGs for a repressor, the consequences of its loss are likely to be more serious. In contrast, conservation of a target gene but not one or more of its multiple parallel-acting activators is unlikely to shut-off gene expression completely and, at the same time, not result in major cost from spurious overexpression. Thus, the differential evolutionary dynamics of repressors and activators might be significant in a closely related group of organisms, especially as they adapt to parasitic niches through genomic stream-lining. Objective tests for the generality of the observations of Hershberg and Margalit and their explanations should be soon possible with reconstruction of TRNs of other major bacterial radiations.

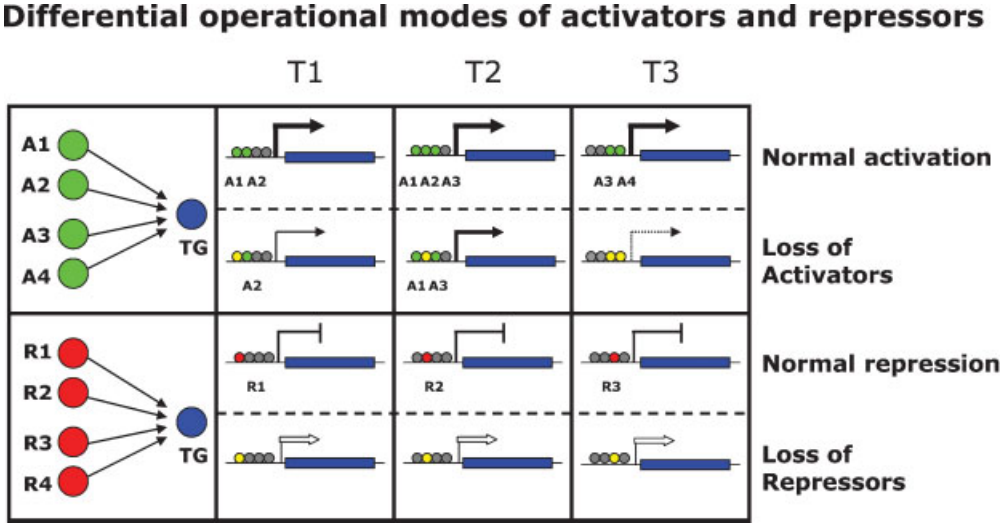

Figure 3. A schematic representation of hypothesized distinct operational modes of activators and repressors at different physiological states of cell. These states are denoted by time points T1 to T3. Activators and repressors are represented by green and red circles, while the target genes are denoted by blue circles or rectangles. It could be easily seen from the figure that there are multiple activators providing parallel inputs to a target gene at any given time. But repressors act in a serial mode providing single input at any time. The circles denoted in grey and yellow correspond to absence and loss of transcription factors, respectively. The gene expression levels, spurious and base-level expressions are indicated respectively by thickness of the arrows, open arrows and dotted arrows.

Acknowledgments

We gratefully acknowledge the intramural research program of the National Institutes of Health for funding our research. We thank M. Madan Babu, Raja Jothi and S. Geetha for useful comments about the manuscript.

References

1. Hershberg R, Margalit H. 2006. Co-evolution of transcription factors and their targets depends on mode of regulation. Genome Biol 7:R62.
2. Salgado H, Santos-Zavaleta A, Gama-Castro S, Peralta-Gil M, Penaloza-Spinola MI, et al. 2006. The comprehensive updated regulatory network of Escherichia coli K-12. BMC Bioinformatics 7:5.
3. Babu MM, Luscombe NM, Aravind L, Gerstein M, Teichmann SA. 2004. Structure and evolution of transcriptional regulatory networks. Curr Opin Struct Biol 14:283–2891.
4. Barabasi AL, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113.
5. Thieffry D, Huerta AM, Perez-Rueda E, Collado-Vides J. 1998. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. Bioessays 20:433–440.
6. Teichmann SA, Babu MM. 2004. Gene regulatory network growth by duplication. Nat Genet 36:492–496.
7. Madan Babu M, Teichmann SA, Aravind L. 2006. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. J Mol Biol 358:614–633.
8. Lozada-Chavez I, Janga SC, Collado-Vides J. 2006. Bacterial regulatory networks are extremely flexible in evolution. Nucleic Acids Res 34:3434–3445.
9. Dekel E, Alon U. 2005. Optimality and evolutionary tuning of the expression level of a protein. Nature 436:588–592.
10. Pavlidis P, Noble WS. 2003. Matrix2png: a utility for visualizing matrix data. Bioinformatics 19:295–296.