“Bacterial regulatory networks are extremely flexible in evolution”

Irma Lozada-Chávez & Julio Collado-Vides

Program of Computational Genomics. Center for Genomic Sciences, UNAM. México.
GRC, July 2006. Correspondence to: ilozada@ccg.unam.mx

REFERENCES

[1] Thieffry *et al* (1998). *BioEssays* 20:433-440.
[2] Teichmann & Babu (2004). *Nat. Gen.* 36:492-496.
[3] Martínez-Antonio & Collado-Vides (2003). *Curr. Opin. Microbiol.* 6:482-489.
[4] http://regulondb.ccg.unam.mx/index.html
[5] http://dbtbs.hgc.jp/
[6] *Yu et al (2004). Genome Res.* 14:1107-1118.
[7] http://www.weizmann.ac.il/mcb/UriAlon/groupNetworkMotifSW.html
[8] de Hoon, *et al* (2004). *Bioinformatics* 20:1453-1454.

INTRODUCTION

Evolution is the result of the variation and selection through time of the components and structure of the organisms, for example, the transcriptional basis of natural variation in microbial species.

Transcriptional regulation plays a prominent role because it controls the cellular processes through activity of proteins and the expression levels of genes for the response to environmental changes, such as nutritional status and several stresses.

An important idea emerging from works of Michael A. Savageau and Stuart A. Kauffman (around 70’s) is that the transcriptional regulation can be viewed as a **complex network** of interactions among diverse types of molecules like proteins, DNA and metabolites [1].

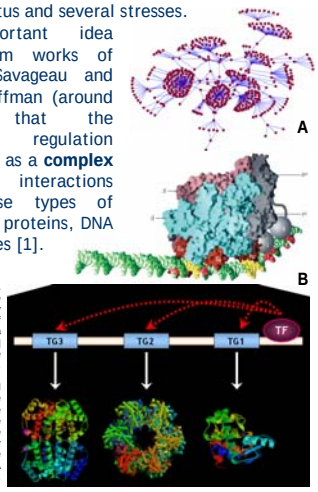


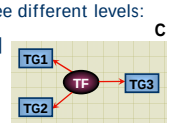
Figure 1.
B-C) Basic unit of the gene regulatory interaction consists of three elements: a **transcription factor [TF]** its DNA binding site and the **target gene(s) [TG]** [2].

A) Transcriptional Regulatory Network in the species is complex because some genes may be regulated by more than one TF and some TFs may control more than one gene through the same DNA binding site(s) [3].

OBJETIVE

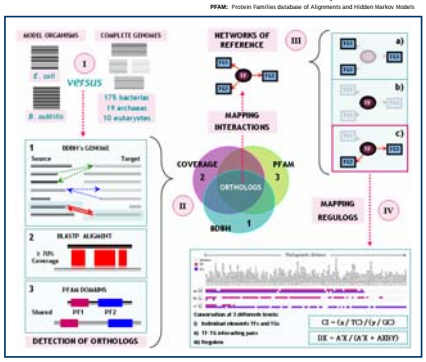
In order to understand the plasticity of **Transcriptional Regulatory Network [TRN]** in bacteria, we studied the conservation of currently known TRNs of the two model organisms *Escherichia coli K12* [4] and *Bacillus subtilis* [5] across complete genomes including **Bacteria**, **Archaea** and **Eukarya** through **REGULOG** [6] principle at three different levels:

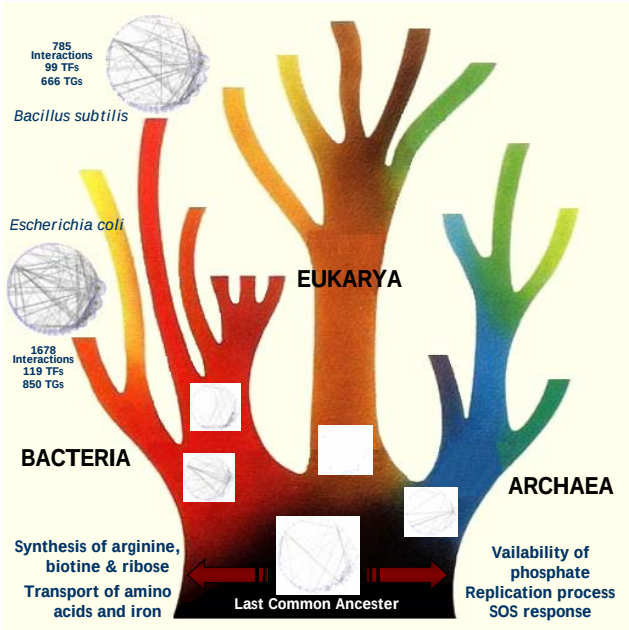
1) Transcription Factors [TF] and **Target Genes** [TG]
2) Pairs of interactions
3) Regulons
(group of all genes regulated by a TF)



METHODOLOGY

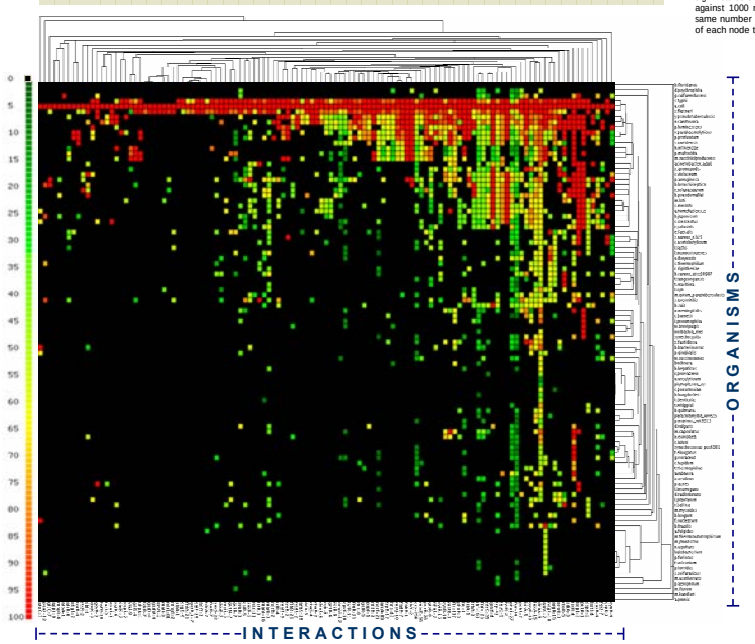
BBHs: Bi-directional Best Hit
BLASTP: Basic Local Alignment Search Tool for Protein
PFAM: Protein Families database of Alignments and Hidden Markov Models





We demonstrate that individual elements, interacting pairs and groups of interactions are not conserved, in fact even in closely related species.

Therefore, transcriptional regulation reflects that is more flexible than genetic component of the organisms to adapt to environmental changes in phenotypic speciation.

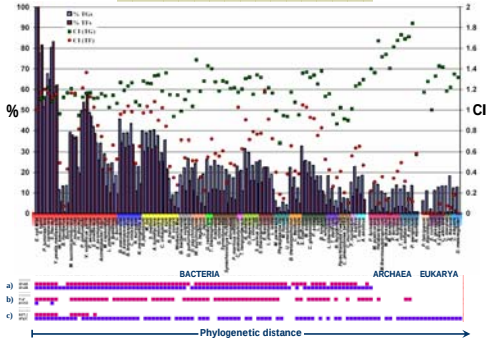


RESULTS

Transcription factors [TFs] evolve much faster than target genes [TGs] across phyla. Hence, TFs could be major players responsible for the evolvability of the TRNs.

Figure 2. Conservation of the components of the TRN (TFs and TGs) across the three cellular domains from *Escherichia coli* K12. In X axis are 110 non-redundant genomes ordered by phylogenetic distance. In Y-axis (to the left) is the percentage of conservation of the elements (TFs and TGs) of the TRNs. CI values (shown to the right on the Y-axis) represent a measure of conservation of the components of the TRN of a genome with respect to the conservation of its genes. Color codes on X-axis represent 13 different phylogenetic clades that correspond to the specific position in the phylogenetic tree of cellular domains.

$$CI_{(TF/TG)} = \frac{\text{(ortologs TF or TG / TFs or TGs of E. coli's TRN)}}{\text{(ortologs of E. coli's proteome / E. coli's proteome)}}$$



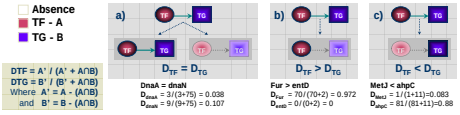
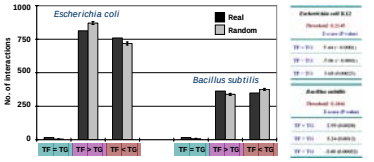


Figure 3. Classification of TF-TG pairs into three different categories. Examples of TF-TG pairs distributed in to different classes based on their co-evolution pattern: a) TFs and TGs co-evolve, b) TF is evolutionarily more conserved than TG and c) TF is less conserved than TG. To evaluate the statistical significance of the conservation of the regulatory interactions in these three different classes, we compared against 1000 randomly constructed regulatory networks for *E. coli* and *B. subtilis* each composed of the same number of interactions as the original TRNs but by switching the edges while maintaining the degree of each node the same as in the known TRN [7].

There is only a small fraction of significantly conserved interactions in bacteria and there is no constraint on the elements of the interaction to co-evolve.



The majority of regulons (interactions) in bacteria are rapidly lost implying a high order flexibility in the TRNs.

Figure 4. Conservation of 1678 pairs of regulatory interactions grouped in 118 regulons of the *E. coli*’s TRN across the 110 complete genomes. For each TF in *E. coli*, we calculated the percentage of total interactions conserved in its regulon across genomes. To represent this distribution we clustered by the extent of TRN and regulon conservation using Centroid Linkage Clustering method with an Uncentered Correlation as distance metric from the Cluster program [8]. **Horizontal conservation** of the TRN reflect the general phylogenetic distribution of Brown et’al (2001) and Yang et’al (2005) for the proteobacteria and archaea, all obligated parasitic and endosymbiotic organisms were grouped together. **Vertical conservation** of the TRN shows the most conserved regulons across bacterial species. We calculated clustering data for 93 regulons of the *B. subtilis*’ TRN (results not shown).