

# Escherichia coli K-12 as a model to understand gene regulation

*Escherichia coli* K-12, a bacterium of the normal flora of humans, represents one of the most important model organisms in biology. This microorganism has been completely sequenced at DNA level, and contains 4319 genes. An important element associated with gene expression in this bacterium corresponds to DNA-binding regulatory proteins or Transcription Factors (TFs). These proteins provide *E. coli*, and in general all the organisms, with the ability to contend against environmental changes, by either blocking (negative) or allowing (positive) gene expression, depending on the environment conditions and metabolic status. In this context and motivated by the massive amount of experimental evidences accumulated during the last decade on this bacterium, we asked for three main questions: what is the repertoire of proteins devoted to regulate gene expression in *E. coli* K-12?; how many genes could be regulated by these proteins?; and how many genes could be associated with alternative regulatory mechanisms?.

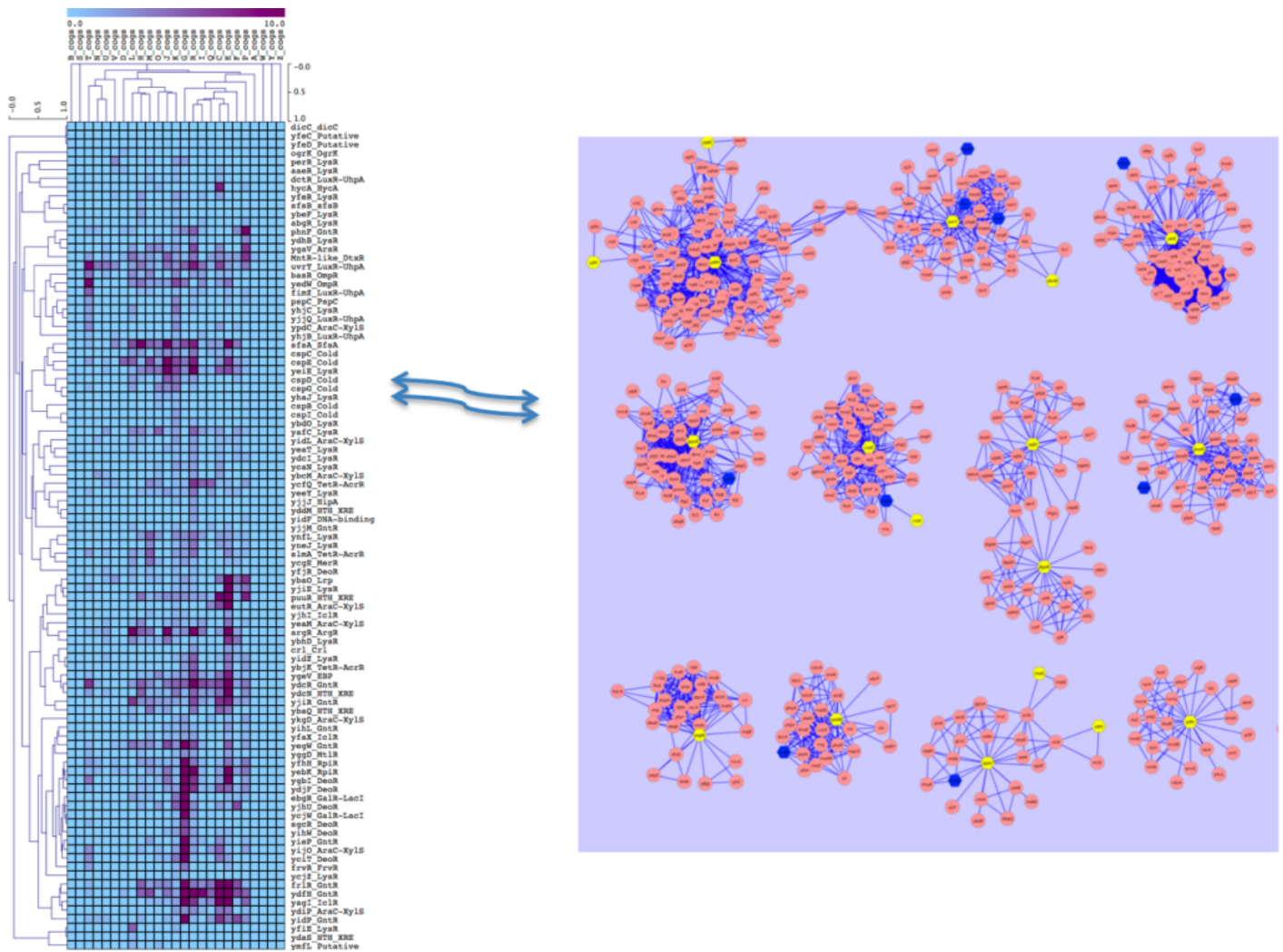


Fig. 1. Identifying functional modules in the regulatory network of *Escherichia coli* K-12.

Therefore, based on bioinformatics analysis, we identified 304 proteins that *E. coli* could use at different conditions to regulate its gene expression, i.e. almost 7% of the total gene repertoire in this bacterium. From this, almost 60% of the regulatory proteins have been experimentally characterized, whereas the other half represents new regulatory proteins. Experimental evidences suggest that the repertoire of well-characterized regulatory proteins control the expression of around the 36% of the total genes in *E. coli*, suggesting that a large number of potential new Regulatory protein- regulated gene interactions are to be discovered. In this regard, we estimated that we are close to the optimum number of TFs in this bacterium, as it has already been estimated by other authors considering the number of interactions experimentally characterized; however, the existence of alternative regulatory mechanisms already described in this organism, such as riboswitches, or DNA supercoiling, are emerging in the last years.

In addition, in this manuscript we identified functional interactions among the *E. coli* regulatory proteins and regulated genes based on diverse approaches: functional linkages among genes which fuse to form a single gene; phylogenetic profiles, i.e. the evolutionary history of one protein in the context of its presence or absence in all the organisms; and the chromosomal association of bacterial genes in operons, a basic structure of gene organization in bacteria. Therefore, taking into account this information, functional groups were identified. These groups include regulatory proteins belonging to the same evolutionary families, regulating genes involved in similar physiological functions, suggesting that in functional and evolutionary terms those clusters are conserved.

We consider that the compilation and analysis of regulatory elements in *E. coli* help us to understand how the regulatory network is organized to respond efficiently to environment challenges. Despite the fact that regulatory proteins are the most extensively used element in regulatory networks, the extended repertoire of other regulatory mechanisms has resulted in a significant increase in the versatility of the network, accurately modulating the organism's gene expression. In addition, we found a large proportion of regulatory proteins regulating similar functions and when they are analyzed together with the uncharacterized regulatory proteins, their coupling are consistent as robust clusters of similar functions. Finally, global regulators, i.e. those regulators devoted to regulate large proportions of genes, were found clustered together, suggesting functional and evolutionary relationships. Altogether, this analysis provides new clues about *E. coli* genetic regulation network that can be expanded to other organisms.

## Publication

[The functional landscape bound to the transcription factors of Escherichia coli K-12.](#)

Pérez-Rueda E, Tenorio-Salgado S, Huerta-Saquero A, Balderas-Martínez YI, Moreno-Hagelsieb

G

*Comput Biol Chem.* 2015 Oct