Toolbox model of evolution of prokaryotic metabolic networks and their regulation

Sergei Maslov[1], Sandeep Krishna[2], Tin Y. Pang[1,3], and Kim Sneppen[2]

***Short Abstract* — It is observed that the number of transcription factors in prokaryotic genomes scales approximately quadratically with the genome size. Here we propose a simple model to explain this observation. The model views a prokaryotic genome as a toolbox composed of functional pathways and interconnected by a network, and considers the adaptation of an organism to a new environment by gaining new pathways and shedding away the useless ones. It suggests that as the genome size gets bigger, the number of genes in newly acquired pathways gets smaller, as the organism can reuse its existing genes. This explains the faster than linear scaling observation between the number of transcription factors and genome size, and we provide a simple toolbox interpretation to simplify our conclusion..**

***Keywords* — Functional genome analysis, horizontal gene transfer, transcriptional regulatory networks**

I. Introduction

THE biological functions of a cell are carried out by the metabolic network and controlled by the regulatory network. The coordination of metabolism in the cell by the regulatory network is very extensive, given the fact that about half of the transcription factors in *Escherichia coli* have a binding site for a small molecule [1], and these transcription factors could be potential regulators of metabolic pathways [2]. This indicates that coordination of metabolic pathways is the major function of the regulatory network. There are empirical evidences that shed light on the evolutionary process between the two networks: a) the number of transcription factors scales faster than linear [3-6] (approximately quadratically [4]) with the total number of proteins in the prokaryotic genomes and b) the distribution of regulons, that is, the out degree distribution of transcription factors in the regulatory pathways has a long tail [7]. This implies that, excluding a few "hub" like transcription factors, all work locally. Here we propose a simple model to explain the above observations based on the co-evolution of bacterial metabolic network.

II. Toolbox view of the metabolic network

We propose to view the collection of enzymes in a organism as tools in a toolbox, where every enzyme is a tool that carries out a task either breaking down large metabolites into smaller ones or assembling simple metabolites into complicated ones. Our model shows that it is the reuse of enzymes that gives rise to the nonlinear scaling. Adapting to a new environmental condition, e.g. presence of a new food source, would require new metabolic pathways to process the new metabolites. The new metabolic pathways can be made up of the newly acquired enzymes / tools and also existing enzymes / tools in the genome / toolbox. In this analogy it is clear to see that the larger the size of the toolbox the higher the probability of reusing the existing tools. In terms of the metabolic network language, we have the larger the number of enzymes, the higher the probability of reusing those enzymes that are already present on a new metabolic pathway. The spirit of the toolbox analogy can be further extended on the whole prokaryotic genome since the metabolic genes constitute a large portion of it. In the sense the larger the size of a genome, the higher probability to have a gene in the genome being reused in a new functional pathway to adapt to changes in the environment, and therefore the lower the number of new genes for that new functional pathway. If we assign a transcription factor to every functional pathway, then we will get a nonlinear scaling between the number of transcription factors and the genome size.

[1]Department of Condensed Matter Physics and Materials Science, Brookhaven National Laboratory, Upton, NY 11973

[2]Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark

[3]Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794-3800

References

[1] Madan Babu M, Teichmann SA (2003) Evolution of transcription factors and the gene regulatory network in Escherichia coli. Nucleic Acids Res 31(4):1234 –1244.

[2] Anantharaman V, Koonin E, Aravind L (2001) Regulatory potential, phyletic distribu-tion and evolution of ancient, intracellular small-molecule-binding domains. J Mol Biol 307(5):1271–1292.

[3] Stover C, et al. (2000) Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. Nature 406(6799):959 – 964.

[4] van Nimwegen E (2003) Scaling laws in the functional content of genomes. Trends Genet 19(9):479 – 484.

[5] Cases I, de Lorenzo V, Ouzounis C (2003) Transcription regulation and environmental adaptation in bacteria. Trends Microbiol 11(6):248 –253.

[6] Konstantinidis K, Tiedje J (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci USA 101(9):3160 –3165.

[7] Thieffry D, Huerta A, 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(5):433–440.