

Toolbox Model of Evolution of Metabolic Pathways on Networks of Arbitrary Topology

Tin Yau Pang^{1,2} Sergei Maslov^{1*} (maslov@bnl.gov)

¹Biology Department, Brookhaven National Laboratory, Upton, , New York

²Department of Physics and Astronomy, Stony Brook University, New York

Project Goals: The biological functioning of a living cell involves coordinated actions of its metabolic and regulatory networks. Metabolic networks are composed of many semi-autonomous functional units – metabolic pathways. These pathways are routinely controlled by dedicated transcription factors, and the activities of individual pathways need to be well coordinated with each other. The project goal is to investigate general principles behind such coordination in prokaryotic genomes. To this end we carry out dynamical and evolutionary modeling of the integrated network encompassing metabolic and regulatory interactions.

It has been previously reported¹ for prokaryotic genomes that the number of Transcription Factors (TFs) is proportional to the *square* of the total number of genes. As a consequence of this trend the fraction of TFs (the so-called "regulatory overhead") is less than 0.5% in small (< 500 genes) bacterial genomes, while in large genomes (~10,000 genes) it can be as high as 10%. We recently proposed² a general explanation of this empirical scaling law and illustrated it using a simple model in which metabolic and regulatory networks co-evolve together. In our model prokaryotic organisms acquire new metabolic functions by the virtue of horizontal gene transfer of entire co-regulated metabolic pathways from a shared gene pool (the "universal metabolic network" or bacterial metabolic pan-genome). This transfer is followed by removal of redundant enzymes and assignment of a dedicated TF regulating the newly acquired pathway³. The whole process can be compared to a homeowner buying sets of tools from a hardware store and later returning duplicate items. We view the full repertoire of metabolic enzymes (or more generally all non-regulatory proteins) encoded in the genome of an organism as its collection of tools. Adapting to a new environmental condition (e.g. learning to utilize a new nutrient source) involves acquiring new tools as well as reusing some of the tools that are already encoded in the genome. As the toolbox of an organism grows larger, it can reuse its existing tools more often and thus needs to acquire fewer new enzymes to master each new functional task. From this argument it follows that, in general, the number of metabolic pathways and their regulators should always scale faster than linearly with the total number of genes in a genome. The empirically observed quadratic scaling between these two numbers can be mathematically derived for a broad range of universal network topologies⁴. Furthermore, the sizes of evolutionary conserved pathways in our model have a long-tailed power-law distribution that agrees with empirical observations. This offers a conceptual explanation for the empirically observed broad distribution of regulon sizes or TFs out-degrees in regulatory networks.

References

1. E van Nimwegen, "Scaling laws in the functional content of genomes", Trends Genet 19, 479-84 2003
2. S Maslov, S Krishna, T Y Pang, K Sneppen, "Toolbox model of evolution of prokaryotic metabolic networks and their regulation", PNAS 106, 9743-9748 2009
3. J Grilli, B Bassetti, S Maslov, and M Cosentino Lagomarsino, "Joint scaling laws in functional and evolutionary categories in prokaryotic genomes", Nucleic Acids Research doi: 10.1093/nar/gkr711 2011
4. TY Pang, S Maslov, "Toolbox model of evolution of metabolic pathways on networks of arbitrary topology" PLoS Comp. Bio 8, e1001137 2011.

Support of this work was provided by the DOE Systems Biology Knowledgebase project "Tools and Models for Integrating Multiple Cellular Networks"