

Evolution of new regulatory functions

Tamar Friedlander^{1,2*}, Roshan Prizak^{2*}, Nicholas H. Barton² and Gašper Tkačik²

Short Abstract — Gene expression is controlled by regulatory proteins interacting specifically with external signals and DNA regulatory sequences. These interactions force the network components to co-evolve to maintain functionality. Yet, existing evolutionary models mostly focus on isolated genetic elements. Here we construct a network model to study the evolutionary expansion of gene regulatory networks via duplication and subsequent specialization. We synthesize a biophysical model of molecular interactions with the evolutionary framework to find the conditions and pathways by which new regulatory functions emerge. We show that specialization is usually slow, but is accelerated by regulatory crosstalk and mutations that promote promiscuous interactions.

Keywords — biological networks, molecular evolution, gene regulation.

I. INTRODUCTION

GENE regulation is flexible and its evolution is thought to be more rapid than the evolution of the coding sequences. The case that we focus on here is the divergence of gene regulation, via expansion of transcription factor (TF) families. Following such expansion, a regulatory function is carried out by a larger number of TFs than before, allowing for additional fine-tuning or for an expansion of the regulatory scope. The main avenue for such expansions are TF duplications. Subsequent specialization of TFs often involves divergence in both their inputs (e.g., ligands) and outputs (regulated genes). Despite its key role, theoretical understanding of TF duplication is still incomplete. Existing models predominantly belong to two categories: gene duplication–differentiation models study sub-functionalization of isolated proteins with no regulatory role; Biophysical models use a thermodynamic description of TF–BS (binding site) interactions accounting for the broad DNA binding repertoire of TFs [1], but disregard gene duplications. Here we synthesize these two frameworks to construct a biophysically realistic description of gene regulatory network evolution.

Acknowledgements: This work was funded by People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement Nr. 291734 (T.F.), ERC grant Nr. 250152 (N.B.) and Austrian Science Fund FWF P28844 (G.T.).

¹The Robert H. Smith Institute of Plant Sciences and Genetics in Agriculture, Faculty of Agriculture, Hebrew University of Jerusalem, P.O.Box 12 Rehovot 7610001, Israel, ²IST Austria, Am Campus 1, Klosterneuburg 3400, Austria. E-mail: tamar.friedlander@mail.huji.ac.il.

II. RESULTS

We assume two TF copies, potentially responding to two signals and regulating two or more downstream genes. A genotype in our model is specified by the regulatory DNA binding site sequences, TF binding preferences and signal sensitivities. Using the thermodynamic model of gene expression [1] we can calculate for each genotype its gene expression in response to the input signal(s). Genotypes evolve via combinations of mutations affecting TF binding preferences, signals sensitivities or binding sites. These mutations are then either fixed or lost depending on their fitness effect, whereas fitness is determined by the network gene expression. This defines a huge, yet finite fitness landscape. Using Markov chain framework, we are able to fully calculate its steady state, dynamics and evolutionary trajectories. Importantly, network functionality in our model is determined not by any particular sequence, but rather by the match or mismatch between sequences of distinct components. This enables us to coarse-grain the huge genotype space into only six “macro-states” based on interaction intensities. Such mapping significantly simplifies the analysis and demonstrates the huge dimensionality reduction between genotype to phenotype. We find two possible evolutionary trajectories to specialization: either going via intermediate configurations of partial specificity or via temporary loss and re-gain of TF specificity.

As TFs and BS should co-evolve, they constrain each other. We find that TF evolution becomes slower and more constrained the more downstream genes it regulates. We propose that mutations that reversibly broaden the TF binding scope (“promiscuity-promoting”) can alleviate these constraints and shorten evolutionary times [2].

III. CONCLUSIONS

The novelty of our work is both conceptual and methodological. While most evolutionary models focus on single genes, here we demonstrate that network evolution is radically different. We develop methodologies to analyze high-dimensional genotype spaces and interpret network phenotypes, that are more broadly applicable [3].

REFERENCES

- [1] Ackers G, Johnson A, Shea M (1982) Quantitative model for gene regulation by lambda phage repressor. *PNAS* 79:1129-1133.
- [2] Pougach, K. et al. (2014) Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. *Nat. Commun.* 5, 4868.
- [3] Friedlander, T. et al. (2017) Evolution of new regulatory functions on biophysically realistic fitness landscapes. *Nat. Commun.* 8, 216.