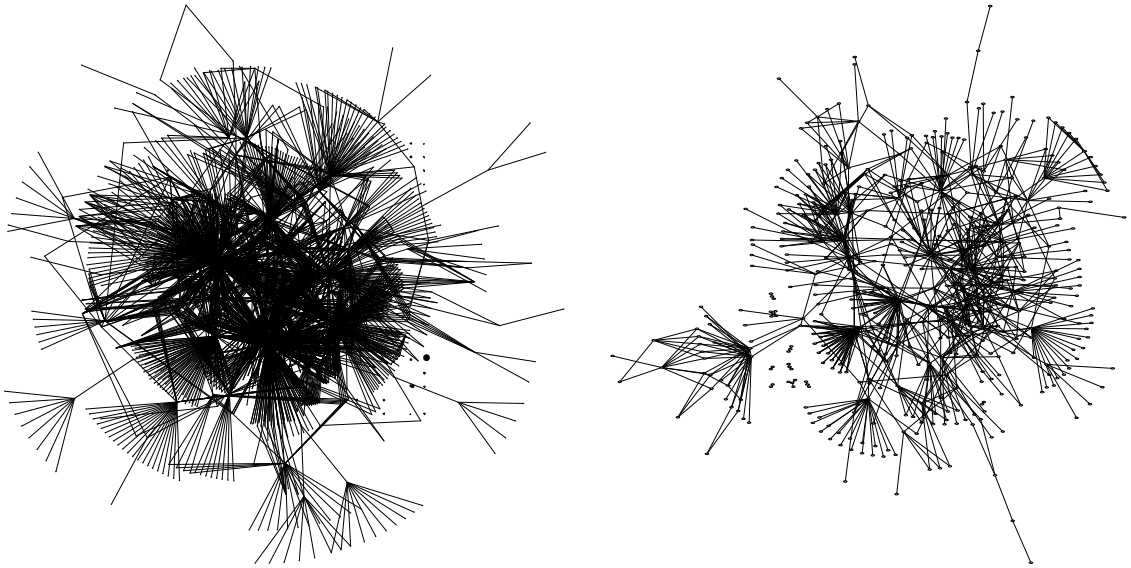
Modeling the Evolution of Gene Regulatory Networks

Joshua W.K. Ho[1,2,3,*], Michael A. Charleston[1,*],

1. School of Information Technologies, University of Sydney, Sydney, Australia
2. School of Biological Sciences, University of Sydney, Sydney, Australia
3. National ICT Australia, Sydney, Australia

*E-mail: joshua@it.usyd.edu.au, mcharleston@it.usyd.edu.au

All biological processes in a cell are tightly controlled by its gene regulatory network (GRN). Since the robustness of a GRN is ultimately determined by its structure, one of the grand challenges in systems biology is to understand how the structures of GRNs evolve. We studied the structure of the established gene regulatory networks of *E. coli* and yeast (Figure 1), using a range of topological features. In comparison with random networks generated by the Erdös-Rényi model and a degree distribution preserving "edge rewiring" algorithm, real GRNs are shown to have degree distribution free, non-random patterns of autoregulators, longest path lengths, clustering coefficients and feed-forward loops (FFLs). All but the largest network components were found to be single-input modules (SIMs) with a universal activator or repressor. Our results indicate that the complexity of GRN structures cannot be entirely captured by a simple scale-free architecture, and this motivated the development of our more biological model.



(a) *E. coli* (b) Yeast

Figure 1: The gene regulatory networks of *E. coli* and yeast. The *E. coli* GRN contains 1306 genes and 2981 interactions while the yeast GRN contains 477 genes and 906 interactions. Despite the differences in size, both networks show similar overall structure and node degree distribution. Both GRNs consist of one large network component with a highly interconnected regulatory core, and a few small disconnected components.

We created a probabilistic GRN growth model, comprising gene duplication, horizontal gene transfer and mutation. Biologically meaningful model parameters were estimated for both

E. coli and yeast GRNs, confirming a high rate of gene duplication events, coupled with preferential duplication of non-regulators. Our model outperforms preferential attachment based models in generating networks, by better capturing the node degree distribution and a range of topological characteristics of real GRNs. Our simulation results support the sub-neofunctionalization hypothesis of duplicate genes in which prolonged neofunctionalization is observed after rapid subfunctionalization [4]. We observed that GRN structures are more likely to be generated if autoregulators are present in the early stages of evolution. Given the estimated low rate of establishing new interactions, duplication of autoregulators provides a simple pathway to establish regulatory links between regulators, which then undergo independent subfunctionalization. This implies that autoregulation fuels the evolution of complex regulatory network structures.

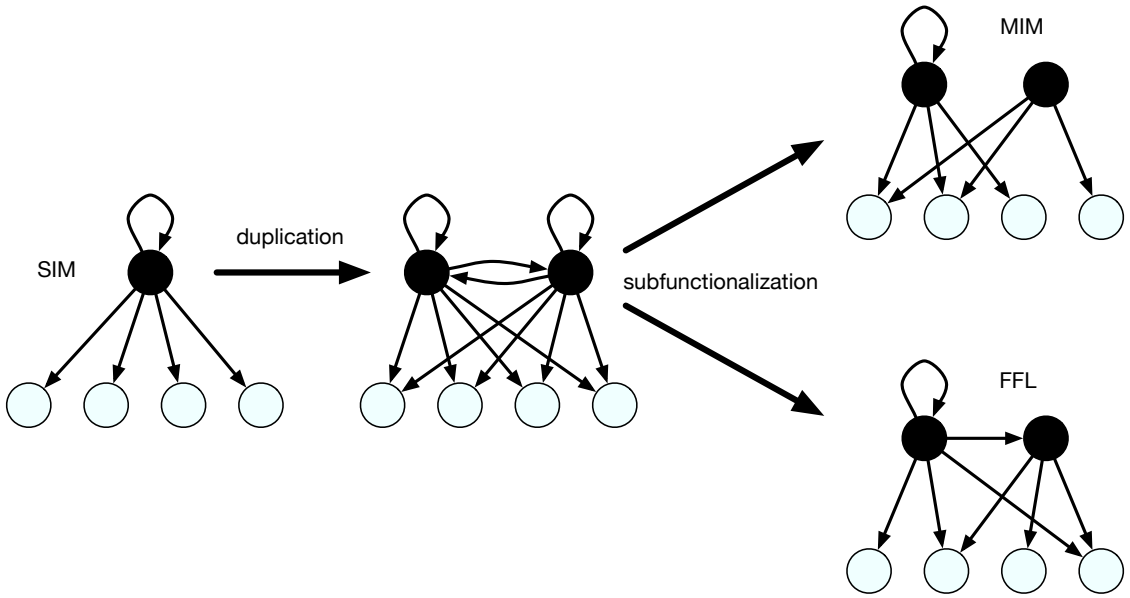


Figure 2: The proposed mechanism for network motif formation. The duplication of an autoregulator in a single-input module can give rise to either a multiple-input module or a set of feed-forward loops.

In light of our computational study and the recent report that autoregulators are mostly proliferated through duplication [3], we propose a simple mechanism in which frequently occurring network motifs such as FFLs and multiple-input modules (MIMs) could arise simply from duplication of the autoregulators in SIMs (Figure 2). If an autoregulator in a SIM is duplicated and subfunctionalized, it can become a MIM or a set of FFLs depending on whether the interactions between the two duplicates are removed. While there have been other mechanisms proposed (e.g. [2]) to explain the overabundance, and presumably the convergent evolution towards, certain GRN motifs [1], we provide a simple mechanism that can explain how such structures could easily arise from known elementary evolutionary processes and natural selection. Our model is implemented in an open-source C++ program called GNLab (http://www.cs.usyd.edu.au/∼mcharles/software/gnlab/).

Acknowledgements

This work is supported by an Australian Postgraduate Award and a National ICT Australia Research Project Award. We thank L.S. Jermiin and members of his laboratory for support and constructive comments. We are grateful to F. Képès for supplying the yeast GRN dataset in text file format.

References

[1] G. C. Conant and A. Wagner. Convergent evolution of gene circuits. *Nat. Genet.*, 34:264–266, 2003.

[2] O. X. Cordero and P. Hogeweg. Feed-forward loop circuits as a side effect of genome evolution. *Mol. Biol. Evol.*, 23:1931–1936, 2006.

[3] M. Cosentino Lagomarsino, P. Jona, B. Bassetti, and H. Isambert. Hierarchy and feedback in the evolution of the *Escherichia coli* transcription network. *Proc. Natl. Acad. Sci. U.S.A.*, 104:5516–5520, 2007.

[4] X. He and J. Zhang. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genet.*, 169:1157–1164, 2005.