How Do Genetic Networks Grow?

Tracing the evolution of *Saccharomyces cerevisiae* using gene duplication events

Aviva Presser[1], Michael Elowitz[2], Roy Kishony[3]

Keywords: protein interaction networks, network motifs, gene duplication, network growth models

1 Introduction.

Protein interaction networks grow larger and more complex as the number of genes in an organism increases. Macroscopic properties of these networks (such as their connectivity distribution) and microscopic properties (such as network motifs [1]) can be used to understand the principle governing their evolutionary growth. A prominent mechanism for network growth is the evolutionary process of gene duplications, which is known to contribute to the formation and development of genomes as we see today. These duplication events can be replication of a single gene or gene segment, or on the other extreme, the duplication of the entire genome. Following such duplication events, there is a transient period that results in either gene loss or functional divergence of the duplicate pair members [2].

We propose a model for network growth incorporating gene duplication and asymmetric divergence of duplicates in the genome, and compare this model to the *S. cerevisiae* protein-interaction network. The model was found to reproduce many of the statistical properties of the observed proteomic network. We then study the motifs in the protein interaction network (small subgraphs occurring in the network in frequencies that are significantly higher than in randomized networks) induced by pairs of duplicated genes. The prevalence of particular network motifs is dependent on the duplication history of the network. From the statistical distribution of these motifs we are able to estimate the rate at which new interactions are formed and old interactions lost in the transient period following a duplication event.

2 Duplication

Kellis et.al. demonstrate [3] that a whole genome duplication has occurred in the history of the *S. cerevisiae* genome. Following the whole genome duplication event in the history of *S. cerevisiae*'s genome, the majority of the newly created genes were lost. However, the genome retains a significant number of duplicate genes. It is the protein interaction patterns of these remaining gene pairs we use to measure parameters pertaining to the transient period of equilibration following the genome duplication.

[1] Division of Engineering and Applied Science, Department of Applied Mathematics, and the Bauer Center for Genomics Research, Harvard University, Cambridge, MA. E-mail: apresser@fas.harvard.edu

[2] Department of Biology, California Institute of Technology, Pasadena, CA. E-mail: melowitz@caltech.edu

[3] Bauer Center for Genomics Research, Harvard University, Cambridge, MA. E-mail: rkishony@cgr.harvard.edu

3 Motifs

We look at the prevalence of network motifs between pairs of duplicate genes in the protein interaction network. Some motifs can appear as a direct result of gene duplication (Figure 1).

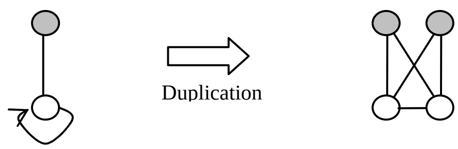

Figure 1: Motif creation by gene duplication

However, not all motifs can be created via gene duplication alone. Many motifs, such as the motif at the end of the process in Figure 2, are only possible if new interactions between the surviving duplicate genes are created, or if old interactions are deleted.

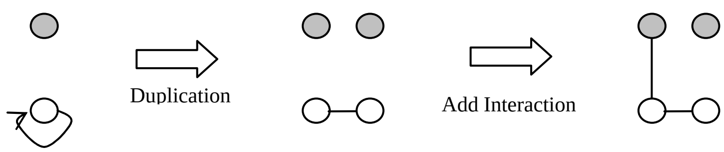

Figure 2: Example of motif not accessible by duplication alone.

We establish the statistical distribution of all possible motifs among pairs of surviving *S. cerevisiae* duplicate genes. Using this distribution we estimate the interaction creation and destruction probabilities (for surviving duplicate pairs) during the transient period following genome duplication.

4 Conclusions

Gene duplication events have played a prominent role in the evolutionary history of genomes, and in the growth of the *S. cerevisiae* genome in particular. A model of gene duplication combined with asymmetric divergence of gene pairs was found to reproduce many of the statistical properties of the *S. cerevisiae* interaction network. We then look at duplication within the genome. From the interaction patterns these pairs exhibit today, we estimate the parameters that governed their functional divergence.

References

[3] Kellis, M., Birren, B.W., and Lander, E. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae.* In: *Nature* vol. 428, pp. 617 – 624.

[1] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chkovskii, D., and Alon, U. 2002. Network Motifs: Simple Building Blocks of Complex Networks. In: *Science*, vol. 298, no. 5594, pp. 824-827.

[2] Wagner, A. 2002. Asymmetric Functional Divergence of Duplicate Genes in Yeast. In: *Molecular Biology and Evolution* vol. 19, pp. 1760-1768 (2002)