

Abstract:

In this grant, we have systematically investigated the integrated networks, which are responsible for the coordination of activity between metabolic pathways in prokaryotes. We have developed several computational tools to analyze the topology of the integrated networks consisting of metabolic, regulatory, and physical interaction networks. The tools are all open-source, and they are available to download from Github, and can be incorporated in the Knowledgebase. Here, we summarize our work as follow.

Understanding the topology of the integrated networks is the first step toward understanding its dynamics and evolution. For Aim 1 of this grant, we have developed a novel algorithm to determine and measure the hierarchical structure of transcriptional regulatory networks [1]. The hierarchy captures the direction of information flow in the network. The algorithm is generally applicable to regulatory networks in prokaryotes, yeast and higher organisms.

Integrated datasets are extremely beneficial in understanding the biology of a system in a compact manner due to the conflation of multiple layers of information. Therefore for Aim 2 of this grant, we have developed several tools and carried out analysis for integrating system-wide genomic information. To make use of the structural data, we have developed DynaSIN for protein-protein interactions networks with various dynamical interfaces [2]. We then examined the association between network topology with phenotypic effects such as gene essentiality. In particular, we have organized *E. coli* and *S. cerevisiae* transcriptional regulatory networks into hierarchies. We then correlated gene phenotypic effects by tinkering with different layers to elucidate which layers were more tolerant to perturbations [3]. In the context of evolution, we also developed a workflow to guide the comparison between different types of biological networks across various species using the concept of rewiring [4], and Furthermore, we have developed CRIT for correlation analysis in systems biology [5].

For Aim 3, we have further investigated the scaling relationship that the number of Transcription Factors (TFs) in a genome is proportional to the square of the total number of genes. We have extended the analysis from transcription factors to various classes of functional categories, and from individual categories to joint distribution [6]. By introducing a new analytical framework, we have generalized the original toolbox model to take into account of metabolic network with arbitrary network topology [7].

Details report

We have carried out our research according to the proposed aims and developed several computational tools. The details are listed as follow.

Aim 1: To develop computational tools for analyzing integrated networks.

An approach for determining and measuring network hierarchy applied to comparing the phosphorylome and the regulome

- Many biological networks naturally form a hierarchy with a preponderance of downward information flow. In this study, we define a score to quantify the degree of hierarchy in a network and develop a simulated-annealing algorithm to maximize the hierarchical score globally over a network. We apply our algorithm to determine the hierarchical structure of the phosphorylome in detail and investigate the correlation between its hierarchy and kinase properties. We also compare it to the regulatory network, finding that the phosphorylome is more hierarchical than the regulome (see Figure 1).

Aim 2: To carry out correlative and topological analysis, using the above tools, in combination with other genomic information

Integration of protein motions with molecular networks reveals different mechanisms for permanent and transient interactions

- The integration of molecular networks with other types of data, such as changing levels of gene expression or protein-structural features, can provide richer information about interactions than the simple node-and-edge representations commonly used in the network community. For example, the mapping of 3D-structural data onto networks enables classification of proteins into singlish- or multi-interface hubs (depending on whether they have >2 interfaces). Similarly, interactions can be classified as permanent or transient, depending on whether their interface is used by only one or by multiple partners. Here, we incorporate an additional dimension into molecular networks: dynamic conformational changes. We parse the entire PDB structural databank for alternate conformations of proteins and map these onto the protein interaction network, to compile a first version of the Dynamic Structural Interaction Network (DynaSIN). We make this network available as a readily downloadable resource file, and we then use it to address a variety of downstream questions. In particular, we show that multi-interface hubs display a greater degree of conformational change than do singlish-interface ones; thus, they show more plasticity which perhaps enables them to utilize more interfaces for interactions. We also find that transient associations involve smaller conformational changes than permanent ones. Although this may appear counterintuitive, it is understandable in the following framework: as proteins involved transient interactions shuttle between interchangeable associations, they interact with domains that are similar to each other and so do not require drastic structural changes for their activity. We provide evidence for this hypothesis through showing that interfaces involved in transient interactions bind fewer classes of domains than those in a control set.

Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators

- Network connectivity has been related to essentiality: Highly connected proteins (hubs) are more important for cell growth and survival. Although this is intuitively reasonable, another way to assess the role of a regulator is to assign it to a level within a "chain-of-command" hierarchy. Here, we analyzed the effects of network rewiring events on transcriptional regulatory hierarchies in two species. First, we superimposed the phenotypic effects of tampering with specific genes and their regulatory connections directly onto the hierarchies. To study second-order effects, which involved changes in the level of regulators within the hierarchy upon deletions or insertions of other regulators or connections, we reconstructed modified hierarchies. We found that rewiring events that affected upper levels had a more marked effect on cell proliferation rate and survival than did those involving lower levels. Moreover, we showed that the hierarchical level and type of change better reflected the phenotypic effect of rewiring than did the number of changes. We also investigated other features connected to the importance of upper-level regulators: In particular, relative to lower-level regulators, upper-level regulators exhibited a greater range of expression values across species, had fewer functionally redundant copies, and had a shorter half-life. Overall, our analysis shows that broadly constructed hierarchies may better reflect the importance of regulators for cell growth than classifications based on the number of connections (hubbiness)

Measuring the Evolutionary Rewiring of Biological Networks

- We have accumulated a large amount of biological network data and expect even more to come. Soon, we anticipate being able to compare many different biological networks as we commonly do for molecular sequences. It has long been believed that many of these networks change, or "rewire", at different rates. It is therefore important to develop a framework to quantify the differences between networks in a unified fashion. We developed such a formalism based on analogy to simple models of sequence evolution, and used it to conduct a systematic study of network rewiring on all the currently available biological networks. We found that, similar to sequences, biological networks show a decreased rate of change at large time divergences, because of saturation in potential substitutions. However, different types of biological networks consistently rewire at different rates. Using comparative genomics and proteomics data, we found a consistent ordering of the rewiring rates: transcription regulatory, phosphorylation regulatory, genetic interaction, miRNA regulatory, protein interaction, and metabolic pathway network, from fast to slow. This ordering was found in all comparisons we did of matched networks between organisms. To gain further intuition on network rewiring, we compared our observed rewirings with those obtained from simulation. We also investigated how readily our formalism could be mapped to other network contexts; in particular, we showed how it could be applied to analyze changes in a range of "commonplace" networks such as family trees, co-authorships and Linux-kernel function dependencies.

The CRIT framework for identifying cross patterns in systems biology and application to chemogenomics

- Biological data is often tabular but finding statistically valid connections between entities in a sequence of tables can be problematic--for example, connecting particular entities in a drug property table to gene properties in a second table, using a third table associating genes with drugs. Here we present an approach (CRIT) to find connections such as

these and show how it can be applied in a variety of genomic contexts including chemogenomics data.

Aim 3: Analytical studies of the toolbox model on graphs of arbitrary topology

Joint scaling laws in functional and evolutionary categories in prokaryotic genomes

- We propose and study a class-expansion/innovation/loss model of genome evolution taking into account biological roles of genes and their constituent domains. In our model, numbers of genes in different functional categories are coupled to each other. For example, an increase in the number of metabolic enzymes in a genome is usually accompanied by addition of new transcription factors regulating these enzymes. Such coupling can be thought of as a proportional 'recipe' for genome composition of the type 'a spoonful of sugar for each egg yolk'. The model jointly reproduces two known empirical laws: the distribution of family sizes and the non-linear scaling of the number of genes in certain functional categories (e.g. transcription factors) with genome size. In addition, it allows us to derive a novel relation between the exponents characterizing these two scaling laws, establishing a direct quantitative connection between evolutionary and functional categories. It predicts that functional categories that grow faster-than-linearly with genome size to be characterized by flatter-than-average family size distributions. This relation is confirmed by our bioinformatics analysis of prokaryotic genomes. This proves that the joint quantitative trends of functional and evolutionary classes can be understood in terms of evolutionary growth with proportional recipes.

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

- In prokaryotic genomes the number of transcriptional regulators is known to be proportional to the square of the total number of protein-coding genes. A toolbox model of evolution was recently proposed to explain this empirical scaling for metabolic enzymes and their regulators. According to its rules, the metabolic network of an organism evolves by horizontal transfer of pathways from other species. These pathways are part of a larger "universal" network formed by the union of all species-specific networks. It remained to be understood, however, how the topological properties of this universal network influence the scaling law of functional content of genomes in the toolbox model. Here we answer this question by first analyzing the scaling properties of the toolbox model on arbitrary tree-like universal networks. We prove that critical branching topology, in which the average number of upstream neighbors of a node is equal to one, is both necessary and sufficient for quadratic scaling. We further generalize the rules of the model to incorporate reactions with multiple substrates/products as well as branched and cyclic metabolic pathways. To achieve its metabolic tasks, the new model employs evolutionary optimized pathways with minimal number of reactions. Numerical simulations of this realistic model on the universal network of all reactions in the KEGG database produced approximately quadratic scaling between the number of regulated pathways and the size of the metabolic network. To quantify the geometrical structure of

individual pathways, we investigated the relationship between their number of reactions, byproducts, intermediate, and feedback metabolites. Our results validate and explain the ubiquitous appearance of the quadratic scaling for a broad spectrum of topologies of underlying universal metabolic networks. They also demonstrate why, in spite of "small-world" topology, real-life metabolic networks are characterized by a broad distribution of pathway lengths and sizes of metabolic regulons in regulatory networks.

Figure 1

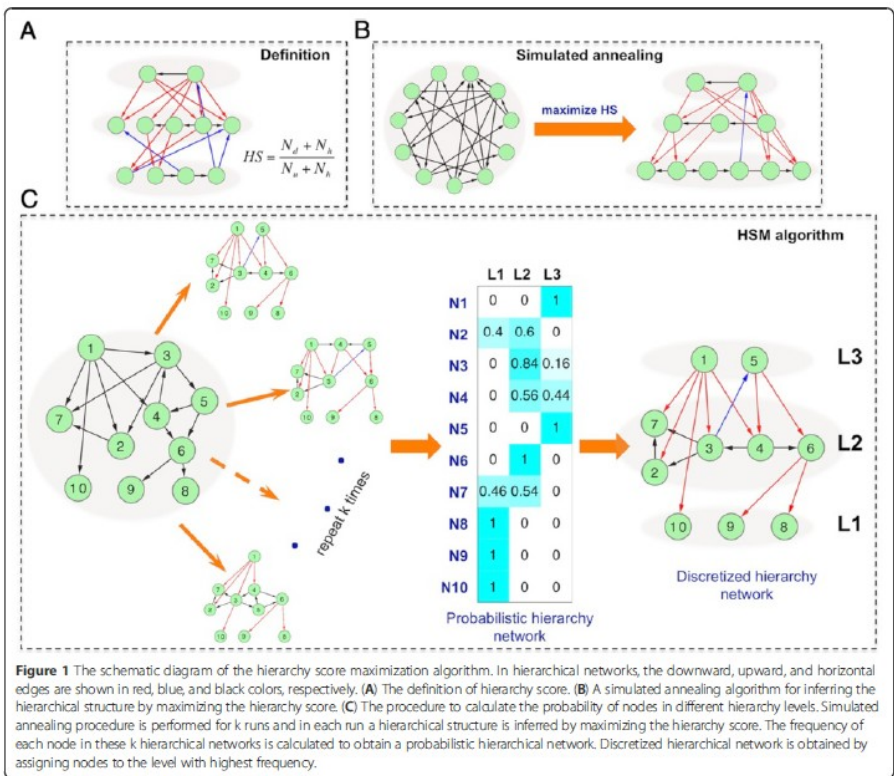


Figure 2

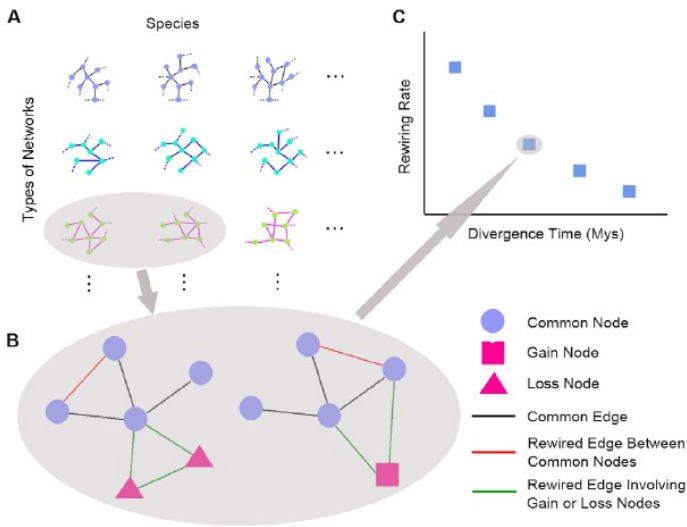


Figure 1. Measuring network rewiring by comparing networks of species pairs. (A) Types of biological networks with currently available data for different species are collected. Selected types of commonplace networks with multiple time-point data are also collected. (B) For each network type, we perform edge rewiring analysis for pairs of species. Three types of nodes are first identified as CNs, GNs and LNs. Four types of rewired edges are then identified and counted including gain/loss edges between CNs (red) and those involving GNs or LNs (green). Rewiring rate from comparing the networks is calculated (see Materials and Methods). (C) Rewiring rate calculated from schematic (B) corresponds to a typical result point.

doi:10.1371/journal.pcbi.1001050.g001

Figure 3

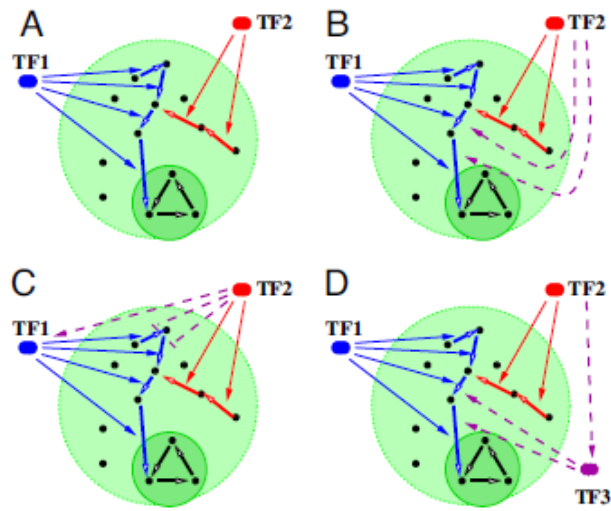


Fig. 2. Schematic diagrams illustrating several possible regulatory network architectures for control of metabolic enzymes/pathways. Four panels correspond to different versions of our model discussed in the article. (A) In the basic model there is no coordination of activity between red and blue metabolic pathways. (B–D) More realistic models include extra regulatory interactions (purple dashed lines) and transcription factors (purple TF3 in D), ensuring that only the part of the blue pathway necessary for utilization of the red nutrient is turned on by the corresponding transcription factor (red TF2).

Reference

- [1] C. Cheng, E. Andrews, K.-K. Yan, M. Ung, D. Wang, and M. Gerstein, "An approach for determining and measuring network hierarchy applied to comparing the phosphorylome and the regulome," *Genome Biol.*, vol. 16, no. 1, p. 63, Mar. 2015.
- [2] N. Bhardwaj, A. Abyzov, D. Clarke, C. Shou, and M. B. Gerstein, "Integration of protein motions with molecular networks reveals different mechanisms for permanent and transient interactions," *Protein Sci.*, vol. 20, no. 10, pp. 1745–1754, Oct. 2011.
- [3] N. Bhardwaj, P. M. Kim, and M. B. Gerstein, "Rewiring of transcriptional regulatory networks: hierarchy, rather than connectivity, better reflects the importance of regulators," *Sci. Signal.*, vol. 3, no. 146, p. ra79, 2010.
- [4] C. Shou, N. Bhardwaj, H. Y. K. Lam, K.-K. Yan, P. M. Kim, M. Snyder, and M. B. Gerstein, "Measuring the Evolutionary Rewiring of Biological Networks," *PLoS Comput Biol*, vol. 7, no. 1, p. e1001050, Jan. 2011.
- [5] T. A. Gianoulis, A. Agarwal, M. Snyder, and M. B. Gerstein, "The CRIT framework for identifying cross patterns in systems biology and application to chemogenomics," *Genome Biol.*, vol. 12, no. 3, p. R32, Mar. 2011.
- [6] J. Grilli, B. Bassetti, S. Maslov, and M. Cosentino Lagomarsino, "Joint scaling laws in functional and evolutionary categories in prokaryotic genomes," *Nucleic Acids Res.*, 2011.
- [7] T. Y. Pang and S. Maslov, "A Toolbox Model of Evolution of Metabolic Pathways on Networks of Arbitrary Topology," *PLoS Comput Biol*, vol. 7, no. 5, p. e1001137, May 2011.