Evolution Enhanced Resilience of Protein Interaction Networks

Jiarui Zhang, Dewu Ding

School of Mathematics and Computer Science
Yichun University
Yichun 336000, China
e-mail: dwding2008@aliyun.com

Abstract—**Most biological functions are presented through protein–protein interaction (PPI) networks. PPI networks show the complex protein-protein interaction relationship within the organism. The generation or destruction of these interaction may lead to changes in biological functions. The latest research results show that the interaction network of species with a higher degree of evolution has higher resilience. Here we explore the resilience changes of a single species during its evolution. We obtain data from public and published websites SNAP. We have proved that no matter what the network structure is, whether it is large or small, when the network fault gradually increases, the resilience gradually decreases. This also indicates that the network with the greatest resilience has a higher degree of evolution.**

Keywords-protein-protein interaction networks; evolution; resilience; network failure

I. INTRODUCTION

The study of life diversity has a long history and a large number of systematic studies have proved that DNA sequence data at the genomic level is constantly evolving [1]. The evolution of DNA and amino acid sequences may rewrite a large number of protein-protein interactions, and generate or destroy the ability of protein interaction networks to express biological functions [2]. On the other hand, organisms adapt to changing environments by adjusting their phenotypes [3], which is also an important source of biodiversity. And most of the phenotypes are the result of interactions between multiple biomolecules and their interactions with the environment.

Therefore, it is necessary to study how the evolution of protein-protein interaction network in a specific organism towards a more adaptive environment. Protein-protein interaction (PPI) network has become an important tool for research; Usually, each node in the network represents a protein, and an edge between two points represents the interaction relationship between them[4].

PPI networks represent complex intracellular protein interactions, and the presence or absence of interactions can lead to biological changes in organisms. In general, not only the mutation of proteins will lead to the deletion of nodes, but also environmental factors or the availability of resources will lead to the destruction of interaction networks. The resulted protein networks decomposition can fundamentally affect the exchange of biological information between proteins in cells.

This paper focuses on the bacterial interaction networks and researches the changes of network resilience when the protein network fails.We construct the node damage model of the protein network to research the change of the resilience when the network fails, and then analyze the change trend of the network resilience in the evolution process.

II. MATERIALS AND METHODS

A. Data sources

The protein interaction networks were obtained from the data published by Zitnik et al [5]. The full interaction networks can be obtained from their website SNAP (http://snap.stanford.edu/tree-of-life/). The protein-protein interaction data includes experimentally supported interactions and human expert-curated interactions rather than computational predictions.

B. Network resilience

Zitnik et al. formally defined network resilience. As an important measure of the topological characteristics of protein networks, network resilience measures the ability of the network to maintain its topological connectivity and measures the anti-interference ability of the network indirectly (the main form of 'interference' is the deletion of network nodes or edges). It measures the speed at which a interaction network crashes when nodes are removed. Zitnik et al. researched the evolution of PPI networks in the whole life tree. The result indicates that the network resilience of biological networks is gradually enhanced in the process of evolution. The greater the network resilience, the stronger the ability of organisms to adapt to the environment.

A interaction network is represented by an undirected graph $G(V, E)$ where G contains N nodes, $V = \{v_1, v_2, ..., v_N\}$, connected by M links, $E = \{(v_i, v_j) | v_i, v_j \in V\}$.

In the original network G, there is a path between any two nodes v_i and v_j. Network resilience is an information theory analysis based on its size, scale and quantity of the connecting components in the network.

When the network is damaged, the connectivity of some nodes in the network G is destroyed. If nodes v_x and v_y are separated in different sub networks G_x and G_y, in other words, there is no path to connect v_x and v_y. These sub networks G_x and G_y are called the connected components of G. The connected components divide G into multiple parts, and the nodes inside the connected component are connected to each other. If v_z has no connection with other nodes in G, v_z is an isolated node. The more a interaction network is damaged, the more isolated nodes will be appear in it.

The improved Shannon diversity was used to calculate the diversity of components as

$$H(G_f) = -\frac{1}{\log N} \sum_{i=1}^{k} p_i \log p_i \qquad (1)$$

It is the entropy of K connected components in G, N is the total number of nodes in the network, $p_i=|C_i|/N$ is the proportion of the number of nodes in the component C_i to the number of nodes in G.

The resilience R of G is then defined as follows:

$$R = 1 - \int_0^1 H(G_f)\,\mathrm{d}f \qquad (2)$$

It is caculated by the accumulation of improved Shannon diversity when f ranges from 0 to 1.

III. RESULT

The recent researches focus on the network resilienec and the evolutionary relationship of species in the life tree. Maddamsetti computed Lenski's long-term evolution experiment data with *Escherichia coli*, and discovered that the evolved PPI networks are more resilient than expected by chance [6].

We study the changes of network resilience when the network is damaged by various failure rates. Generally, if the faulty protein network consists of a majority of large components and a few small components, it is considered that the network has great resilience. On the contrary, if the protein network is divided into many small compone-nts, the network resilience is considered to be low.

Here, network failure is achieved by removing different numbers of nodes, which represent attacks or general node failures. When network nodes are damaged, the edges connected to the nodes will be removed.

We use failure rates r to indicate the degree of network damage. The value range of r is 0%, 1%, 2%, and 100%. When r equals 0, it means that the network is connected without damage. When r equals 1%, it indicates that 1% of nodes in the network are damaged and will be removed from the network.By analogy, when r equals 100%, it means that all nodes in the interactome are damaged and the network does not contain any nodes.

When computing the network failure rate r, randomly remove $r{\times}N$ nodes, and the number of nodes in the network become $N{-}r{\times}N$. Then calculate the resilience of the remaining network. Since the resilience is the rate of network fragmentation calculated by randomly removing nodes, the elastic value is not a constant value. Repeat the calculation for 100 times, that is, for each network failure rate, 100 resilience will be obtained. The 100 resilience of each failure rate r are plotted in the figure, and the resilience of all possible failure rate of r are marked from 0%, 1%, 2%,..., 100%. Then we can get the resilience curve. The curve composed of resilience under all failure rate is called resilience curve.

We focus on PPI networks for three species: *Treponema pallidum*, *Bacillus subtilis* 168 and *Burkholderia pseudomallei* , and draw the resilience curve (Fig.1). In the process of node removal, the resilience decreases gradually and monotonically until 0.
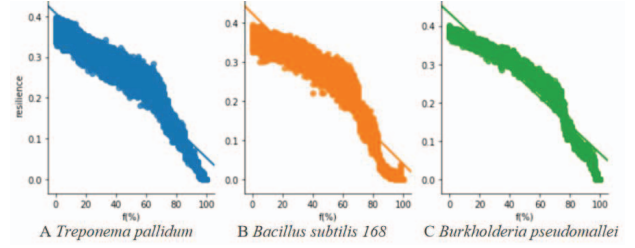

A *Treponema pallidum*   B *Bacillus subtilis 168*   C *Burkholderia pseudomallei*

Figure 1. Resilience change curve. The abscissa is the network failure rate p(0%-100%), and the ordinate is the resilience. Draw all resilience values in the figure with scattered points and fit them into straight lines.

In order to put these results in a better context, we performed a survey of resilience of all bacterial species mentioned in Zitnik pape，and calculated the falling slope of each resilience curve, then used violin chart to count data range (Fig. 2). The slopes of bacteria's resilience curves are mostly between 0.9 and 1.0. When the network failure increases from 0% to 100% gradually, the network resilience decreases significantly. In general, a higher negative slope indicates that deleting nodes produces lower resilience.
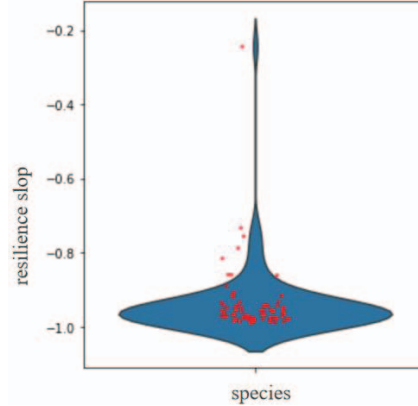

Figure 2. Violin statistics on elastic change curvature. We calculated the slopes of the resilience curves of all common bacteria in zitnik's paper and depicted them with a violin plot.

IV. CONCLUSION

In order to study the changing trend of the resilience of PPI networks in evolution, we systematically calculated the resilience of PPI networks of all bacteria in Zitnik's paper when networks are damaged over all possible failure. We discovered that there exist trends of decreasing resilience in interaction networks when all bacteria encountered network failure. In general, no matter the size of the network, the resilience is gradually reduced when the network suffers from a gradual increase failure. This finding also indicates that the structure of PPI network is formed by evolving towards the direction of increasing resilience, in other words, the evolution enhances the resilience of protein interaction networks.

REFERENCES

[1] Z Yang, "A Space-Time Process Model for the Evolution of DNA Sequences," Genetics, vol. 139, Feb. 2001, pp. 993-1005, doi:10.1093/genetics/139.2.993.

[2] B P Dalrymple , K Kongsuwan, G Wijffels, N E Dixon, P A Jennings, "A Universal Protein–Protein Interaction Motif in the Eubacterial DNA Replication and Repair Systems," PNAS, vol. 98, Sep. 2001, pp. 11627-11632, doi:10.1073/pnas.191384398.

[3] A Charmantier, R H McCleery, L R Cole, C Perrins, L E B Kruuk, B C Sheldon, "Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population," Science, vol. 320, May. 2008, pp. 800-803, doi:10.1126/science.1157174.

[4] J B Pereira-Leal, A J Enright, C A Ouzounis, "Detection of Functional Modules from Protein Interaction Networks," Proteins, vol. 54, Jan. 2004, pp. 49-57, doi:10.1002/prot.10505.

[5] M Zitnika, R Sosica, M W Feldman, J Leskovec, "Evolution of Resilience Inprotein Interactomes across the Tree of Life," PNAS, vol. 116, Feb. 2019, pp. 4426-4433, doi:10.1073/pnas.1818013116.

[6] R Maddamsetti, "Selection Maintains Protein Interactome Resilience in the Long-Term Evolution Experiment with Escherichia coli," Genome Biol Evol, vol. 13, Jun. 2021, pp. 1-7, doi:10.1093/gbe/evab074.