Descriptive analysis of *Arabidopsis thaliana* L. protein-protein interaction network: revealing of key members.

Sergey Pesyak

"DokaGene Technologies" LLC , 58 Moskovskaya Street, Rogachevo, Dmitrov District, Moscow Region 141880, taoekk@gmail.com

Ekaterina Snigir

"TransGene" LLC, 58 Moscovskaya Str., Rogachevo, Dmitrov district, Moscow region, Russia, 141880

t.suprunova@dokagene.ru

The protein-protein interactions networks are in the hearts of various biological processes, including signaling, control of homeostasis, stress responses, growth and development. Such processes are the best known for animals and human, but for plants such studies can reveal much important knowledge about mechanisms of physiological processes. With the development of methods of systemic biology, all the more important become more detailed data and the extraction of biologically significant regularities from them. For this, it is necessary to use methods and algorithms for analyzing topology and graph structures. This is especially important when identifying the critical nodes of the resulting networks, determining in their structure the most significant proteins for a given network.

We analyzed the Arabidopsis thaliana L protein-protein interaction network. The data were collected by the TAIR consortium participants and processed using the programming language R and some packages, including statnet, igraph, dplyr, tidyr, intergraph. Graph structure analysis revealed that it consists of many components, including 8 ordinary vertices without edges, 80 subgraphs with two vertices, and so on to the maximal one. For the purposes of our analysis, which required completely connected graph, we isolate the maximal subgraph into a separate structure for subsequently analyzing.

The size of the selected subgraph was 763 vertex and 1393 edges, with density 0.0024, diameter 19, reciprocity 0.4034458 (almost half of the interactions between proteins are reciprocal), transitivity 0.091 (9% of the triads of proteins are completely connected). As the

triads connections become more complicated, their number decreases linearly, except for two cases: an anomalously small number of type 030C and an anomalously large number of type 201.

For the determination of protein "importance" in our network we used specialized algorithms including PageRank, Hubs, Authority etc. We isolated groups of proteins with the greatest number of links (node degree), as well as the proteins closest to all the actors in the network (closeness centrality) and proteins which play an important role in the information transmitting in our network (betweenness centrality). PageRank, Hubs and Authority showed similar results. The analysis for the presence of bridges (junction points) revealed a large number of proteins, the removal of which destroys the graph connectivity and increases the amounts of components.

A 2-degree clique was most prevalent, the maximal size of the clique was six proteins. Four of six maximal clicks have 5 identical proteins (AT5G39760, AT4G24660, AT1G14440, AT1G75240, AT5G15210). K-cores algorithm showed cores with degrees from one to eight (from 314 to 9 vertices).The study of the results of this analysis revealed several clusters of closely related proteins.

Thus, according to the results of our work, we can conclude structural studying of protein-protein interaction network makes possible to identify the most important for proteins and their groups.

This research was supported by the grant of the Ministry of Education and Science of the Russian Federation (No 2016-220-05-163).

1. A. Aderhold et. al. (2016) Biological Network Inference at Multiple Scales: From Gene Regulation to Species Interactions, In: Pattern Recognition in Computational Molecular Biology: Techniques and Approaches, M. Elloumi et. al. (Eds.), 525–554 (Wiley).

2. A. Chakraborty et. al. (2013) Link Analysis, In: Practical Graph Mining with R, N. F. Samatova (Ed.), 75–134 (CRC Press).

3. S. Fortunato, A. Flammini (2007) Random walks on directed networks: The case of PageRank. *International Journal of Bifurcation and Chaos*,17:2343–2353.

4. Y. Zhang et. al. (2010) Plant Protein-Protein Interaction Network and Interactome, *Current Genomics*, 11:40–46.