

Evolutionary analysis of gene networks with Orthoweb software

Mustafin Z.^{1,2*}, Mukhin A.¹, Anikin D.², Kazantsev F.^{1,2}, Lashin S.^{1,2}

¹ *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

² *Novosibirsk State University, Novosibirsk, Russia*

* *mustafinzs@bionet.nsc.ru*

Key words: phylostratigraphic analysis, gene networks, web-service

Motivation and Aim: Phylostratigraphic analysis is a new trend in evolutionary analysis, aimed to determine the time of gene origin [1]. The method based on analysis of taxonomic classification of genes and their orthologs and finding the common taxon. With such method of sequences analysis, as dN/dS ratio [2], it becomes possible to detect the time of gene's origin and the type of selection affecting to it. The methods of such analysis, implemented in different software packages (myTAI [3], phylostratr [4], Orthoscape [5], etc.), require the user to know the programming languages (mostly R) or how to use another software, like Cytoscape [6]. Most methods also aimed to analyze the gene lists, but it is interesting to look not only to genes, but also to their interactions and possible clusters in gene networks.

Methods and Algorithms: Orthoweb software has been developed in the Java programming language. To develop the server part, we used the Spring framework; VueJS and Webix frameworks have been used for the client part. Networks visualizing is implemented with cytoscape.min.js library. The document-oriented database MongoDB is used to store the data from KEGG database (taxa, list of orthologous, coding sequences, etc.) and intermediate analysis results. It is significantly increases the speed of future work with the same data.

To store the result of analysis, we have developed the PostgreSQL database. Using REST API, the user can retrieve the data from previous analysis of gene with exactly same parameters.

Results and Discussion: We present a web-based service Orthoweb, which allows to analyze both gene sets and gene networks, using only web browser. It is possible to import a gene network to Orthoweb (from one of two external databases, from manually generated file or from automatically generated file by, for example, STRING [7]). Orthoweb visualizes the results of analysis by coloring and sizing the nodes and their borders, and additionally upload the gene ontology terms, associated with genes. Orthoweb allows interacting with nodes and edges and saving the result. Orthoweb also allows to get the results of analysis for model organisms using API of PostgreSQL database to get the data without analysis.

Availability: The web-service is available on <https://orthoweb.sysbio.cytogen.ru>.

Acknowledgements: This study was supported by the Budgetary Project FWNR-2022-0020 and by RFBR No. 20-04-00885.

References

1. Domazet-Lošo T., Brajković J., Tautz D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* 2007;23(11):533-539. doi: 10.1016/j.tig.2007.08.014.
2. Kryazhimskiy S., Plotkin J.B. The population genetics of dN/dS. *PLoS Genet.* 2008;4(12). doi: 10.1371/journal.pgen.1000304.

3. Drost H.-G., Gabel A., Liu J., Quint M., Grosse I. myTAI: evolutionary transcriptomics with R. *Bioinformatics*. 2018;34(9):1589-1590. doi: 10.1093/bioinformatics/btx835.
4. Arendsee Z., Li J., Singh U., Seetharam A., Dorman K., Wurtele E.S. Phylostratr: A framework for phylostratigraphy. *bioRxiv*. 2018;360164. doi: 10.1101/360164.
5. Mustafin Z.S., Lashin S.A., Matushkin Y.G., Gunbin K.V., Afonnikov D.A. Orthoscape: a cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles. *BMC Bioinformatics*. 2017;18(S1):1-9. doi: 10.1186/s12859-016-1427-5.
6. Shannon P. et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504. doi: 10.1101/gr.1239303.
7. Szklarczyk D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-D613. doi: 10.1093/nar/gky1131.