

This website is free and open to all users. Upload and analysis of sensitive personal information requires a login.

webLSA: integrated local similarity analysis and network visualization of biological series data

Yunhui Xiong^{1,*}, Yang Li^{1,*}, Haitong She^{2,*}, Shuaishuai Xu³, Xuanqi Chen², Shengwei Hou³, Li Charlie Xia^{2,#}

{¹School of Mathematics and ²School of Software Engineering}, South China University of Technology, Guangzhou 510641, China

³Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen, 518055, China

* These authors contributed equally to this work.

Corresponding author: Li Charlie Xia: lcxia@scut.edu.cn.

Keywords: Local Similarity Analysis, Time Series Data, Network Visualization

Background

Identifying ecological and biochemical associations in time, spatial and other series of biomolecules from genome, transcriptome and proteome is routinely required in current bioinformatics. Analyzing these co-occurrence and co-expression patterns are fundamental to reveal the underlying biological relations (1,2). Among the computational methods developed for such tasks, Local Similarity Analysis (LSA) has gained significant audience, for its unique ability to captures local and potentially time-delayed (or space-shifted) associations that cannot otherwise be identified with ordinary correlation analysis (3).

LSA has been actively developed (4-7), vigorously validated (2), independently top benchmarked (3), and widely applied with high impacts (1,8-16). Its milestone papers (1,4-7,17) received more than 1200 citations to date. A Python package (eLSA) (4) is freely available for practitioners to conduct LSA, yet feature requests for web-based workflow and integrated network visualization were consistently received. As a response, we present webLSA, an integrated local similarity analysis and network visualization service dedicated to biological series data, with the hope to make LSA more accessible to a wider community.

Methods

The backend of webLSA is based on the same Local Similarity Analysis algorithm that has been developed, improved, and extensively validated since 2006 (4-7). The statistical significance estimation in webLSA employed the approximation theory developed in 2013 (6). webLSA made many improvements including: a) multi-threading acceleration in the backend taking advantage of multi-core CPUs; b) leveraging the Galaxy server (18) for frontend, enabling community-tested and user-friendly data and workflow management; and c) achieving on-the-fly and interactive massive network visualization using the state-of-art Graphistry APIs (19).

Input/Output

webLSA accepts a user uploaded plain txt table containing possibly replicated series data. It also requires the user to submit a set of parameters on the run request form. The front and back ends of webLSA are decoupled so that a user can submit multiple jobs meanwhile monitor the status. The output is a plain txt table containing the local associations and their statistics, p-values, false discovery rates, and other relevant information. The inputs and outputs follow the specification of the widely used eLSA tool, and are explained in the numerous hints, helps and demos available on the webserver. The output is visualizable on webLSA as an interactive network, for intuitive exploration, pruning, annotation and export to standard graph file formats.

Applications

We had a demo page on webLSA tutoring the LSA of a marine microbiome series consisting of 72-hour bi-hourly data of thousands of amplicon sequence variants, with detailed discussions of the identified associations and their ecological implications.

References

1. Steele, J.A., Countway, P.D., Xia, L., Vigil, P.D., Beman, J.M., Kim, D.Y., Chow, C.-E.T., Sachdeva, R., Jones, A.C., Schwalbach, M.S. *et al.* (2011) Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *The ISME Journal*, **5**, 1414-1425.
2. Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J. *et al.* (2016) Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME Journal*, **10**, 1669-1681.
3. Ai, D., Chen, L., Xie, J., Cheng, L., Zhang, F., Luan, Y., Li, Y., Hou, S., Sun, F. and Xia, L.C. (2023) Identifying local associations in biological time series: algorithms, statistical significance, and applications. *Briefings in Bioinformatics*, **24**, bbad390.
4. Xia, L.C., Steele, J.A., Cram, J.A., Cardon, Z.G., Simmons, S.L., Vallino, J.J., Fuhrman, J.A. and Sun, F. (2011) Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates. *BMC Systems Biology*, **5**, 1-12.
5. Ruan, Q., Dutta, D., Schwalbach, M.S., Steele, J.A., Fuhrman, J.A. and Sun, F. (2006) Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics*, **22**, 2532-2538.
6. Xia, L.C., Ai, D., Cram, J., Fuhrman, J.A. and Sun, F. (2013) Efficient statistical significance approximation for local similarity analysis of high-throughput time series data. *Bioinformatics*, **29**, 230-237.
7. Xia, L.C., Ai, D., Cram, J.A., Liang, X., Fuhrman, J.A. and Sun, F. (2015) Statistical significance approximation in local trend analysis of high-throughput time-series data using the theory of Markov chains. *BMC Bioinformatics*, **16**, 301.
8. Bromberg, J.S., Fricke, W.F., Brinkman, C.C., Simon, T. and Mongodin, E.F. (2015) Microbiota—implications for immunity and transplantation. *Nature Reviews. Nephrology*, **11**, 342-353.
9. Chen, L., Collij, V., Jaeger, M., van den Munckhof, I.C.L., Vich Vila, A., Kurilshikov, A., Gacesa, R., Sinha, T., Oosting, M., Joosten, L.A.B. *et al.* (2020) Gut microbial co-abundance networks show specificity in inflammatory bowel disease and obesity. *Nature Communications*, **11**, 4018.
10. Fuhrman, J.A., Cram, J.A. and Needham, D.M. (2015) Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology*, **13**, 133-146.
11. Grady, K.L., Sorensen, J.W., Stopnisek, N., Guittar, J. and Shade, A. (2019) Assembly and seasonality of core phyllosphere microbiota on perennial biofuel crops. *Nature Communications*, **10**, 4135.
12. Needham, D.M. and Fuhrman, J.A. (2016) Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nature Microbiology*, **1**, 1-7.
13. Needham, D.M., Poirier, C., Bachy, C., George, E.E., Wilken, S., Yung, C.C.M., Limardo, A.J., Morando, M., Sudek, L., Malmstrom, R.R. *et al.* (2022) The microbiome of a bacterivorous marine choanoflagellate contains a

- resource-demanding obligate bacterial associate. *Nature Microbiology*, **7**, 1466-1479.
14. Roux, S., Chan, L.-K., Egan, R., Malmstrom, R.R., McMahon, K.D. and Sullivan, M.B. (2017) Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. *Nature Communications*, **8**, 858.
 15. Sun, J., Liao, X.-P., D'Souza, A.W., Boolchandani, M., Li, S.-H., Cheng, K., Luis Martínez, J., Li, L., Feng, Y.-J., Fang, L.-X. *et al.* (2020) Environmental remodeling of human gut microbiota and antibiotic resistome in livestock farms. *Nature Communications*, **11**, 1427.
 16. Weiss, A.S., Niedermeier, L.S., von Stempel, A., Burrichter, A.G., Ring, D., Meng, C., Kleigrew, K., Lincetto, C., Hübner, J. and Stecher, B. (2023) Nutritional and host environments determine community ecology and keystone species in a synthetic gut bacterial community. *Nature Communications*, **14**, 4780.
 17. Ai, D., Li, X., Pan, H., Chen, J., Cram, J.A. and Xia, L.C. (2019) Explore mediated co-varying dynamics in microbial community using integrated local similarity and liquid association analysis. *BMC Genomics*, **20**, 185.
 18. Graphistry: <https://github.com/graphistry/pygraphistry>.
 19. Galaxy: <https://github.com/galaxyproject/galaxy>.