How much can we learn about ancient cells from sequence analysis? New metrics on an old problem.

*Shawn E McGlynn[1], Sarah J Berkemer[2,3]

1. Earth-Life Science Institute, Tokyo Institute of Technology, 2. Bioinformatics Group, Department of Computer Science, Interdisciplinary Center for Bioinformatics, University of Leipzig, 3. Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

A longstanding goal is to apply molecular phylogenetics to understanding ancient physiological and evolutionary states (*e.g.* [1]). With the current explosion of molecular sequencing data (e.g. [2]), it is a good time to consider how far back we can peer with the comparative molecular lens, and ask if we can understand life in its nascent years. A standard test for inferring ancient genes is to look for conservation in microbial genomes, and to analyze phylogenetic branch positions and determine if a gene separates the archaea and the bacteria. However, many phylogenetic trees are cluttered by phenomena such as horizontal gene transfer and non-orthologous displacement [3], making inferences difficult.

In this presentation, we will give an overview of recent work in this area [4, 5], and also present new analyses which help identify ancient proteins. Using the COG database, we built phylogenetic trees for all protein families and analyzed the number of interdomain gene transfer events for each family. We find that proteins assigned to COGs exhibit widely variable amounts of interdomain gene transfer. By using distance matrices which relate *intra*-domain sequence similarity to *inter*-domain similarity, we find that protein families exhibiting numerous inter domain gene transfer events are also most self-similar between domains. Integrating this observation with previous analyses, ancient proteins appear to have large branch lengths separating the domains, but recent proteins are more bushy in tree shape. These findings will be discussed in the context of inferring the characteristics of the most ancient cells. Although this work provides new dimensions to analyze protein families with, more work is needed to definitively identify the proteins in the last common ancestor of the bacteria and archaea.

REFERENCES CITED

1. Woese, C.R.: Bacterial evolution. Microbiol. Rev. 51, 221–271 (1987)
2. Castelle, C.J., Banfield, J.F.: Major New Microbial Groups Expand Diversity and Alter our Understanding of the Tree of Life. Cell. 172, 1181–1197 (2018). doi:10.1016/j.cell.2018.02.016
3. Koonin, E.V.: Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat. Rev. Microbiol. 1, 127–136 (2003). doi:10.1038/nrmicro751
4. Forterre, P.: The universal tree of life: an update. Front. Microbiol. 6, (2015). doi:10.3389/fmicb.2015.00717
5. Weiss, M.C., Sousa, F.L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., Martin, W.F.: The physiology and habitat of the last universal common ancestor. Nat. Microbiol. 1, 16116 (2016). doi:10.1038/nmicrobiol.2016.116

Keywords: LUCA, Early life, Phylogenetics , Bioinformatics

©2019. Japan Geoscience Union. All Right Reserved.