

Molecular Clocks Constrained by Horizontal Gene Transfers Predict an Archaeal Common Ancestor ~3.9 Ga, Coincident with the Proposed Late Heavy Bombardment, and Consistent with the Hypothesis of a Thermophilic Bottleneck for Early Life.

G.P. Fournier¹

¹Department of Earth, Atmospheric, and Planetary Science, Massachusetts Institute of Technology, 54-1016, 77 Massachusetts Ave, Cambridge MA 02139; g4nrier@mit.edu

Introduction: The use of molecular clocks for dating early events in microbial evolution is an ongoing challenge for paleogenomics, as, generally, most ancient microbial lineages lack diagnostic fossil records for the absolute dating of nodes on phylogenetic trees. Therefore, the confidence in previous clock-based estimates for the divergence times of major microbial lineages rests entirely upon the confidence in the unconstrained rate models used, which are sensitive to lineage-specific model violations, and cannot be internally cross-validated.

Here, we attempt to surmount these obstacles by using horizontal gene transfer (HGT) events from archaeal lineages to clades with established fossil records and more reliable calibration dates. This propagates these age constraints of recipient nodes to the ancestors of donor lineages within the archaeal phylogeny, improving the precision of molecular clock models. We present the results of this novel method, which we call TARDIS (Transfer Assisted Rate and Divergence time InferenceS).

Critical to this analysis is the HGT of three SMC complex proteins from methanogens to the ancestor lineage of Cyanobacteria. While cyanobacterial SMC proteins were previously detected as being of archaeal origin [1,2], in-depth phylogenetic analysis shows these transfers likely occurred from the clade containing group II methanogens, Halobacteriales, and Archaeoglobales (G2HA). Since previous molecular clock estimates using fossil and geochemical calibration have placed a high probability on the cyanobacterial ancestor predating 2.5 Ga [3], TARDIS constrains the G2HA donor group to be older still, diverging ~2.9 Ga.

Additional HGTs between archaeal groups and Eukarya, green algae, and cellulolytic bacteria provide additional age calibrations across several archaeal lineages. Secondly, additional HGT events between archaea propagate these constraints to additional nodes within the phylogeny.

Using TARDIS, we show that the most likely age for the last common ancestor of Archaea is ~3.9 Ga, strikingly close to the end of the proposed “Late Heavy Bombardment” series of impact events. This provides independent support for the “bottleneck” hypothesis of early life evolution [4], and is consistent with previous work in ancestral sequence reconstruction predicting

hyperthermophilic Domain ancestors, but a mesophilic Last Universal Common Ancestor [5]. In this model, early life diversified some time earlier than 3.9 Ga, but only hyperthermophiles survived major near-sterilizing impact events, presumably in deep-sea hydrothermal systems. These surviving lineages gave rise to the bacterial and archaeal Domain ancestors. This hypothesis explains the shape, depth, and ecological distributions observed on microbial phylogenetic trees. Expanding the HGT dataset and the quality of cross-calibration constraints will likely further increase the precision of these date estimates.

References: [1] J. Soppa (2001) *Gene.*, 278, 253-64. [2] N. Cobbe & M.M. Heck (2004) *Mol Biol Evol.*, 21, 332-47. [3] B.E. Schirmer et al. (2013) *Proc Natl Acad Sci USA.* 110, 1791-6. [4] M. Gogarten-Boekels et al. (1995) *Orig Life Evol Biosph.* 25, 251-64. [5] M. Groussin et al. (2013) *Biol Lett.* 9, 20130608.