Molecular palaeobiological insights into LUCA and eukaryogenesis

Philip, Donoghue

Phil.donoghue@bristol.ac.uk

The timing and nature of events deep in evolutionary is challenging, not least since it has left an incomplete and incoherent fossil record. Indeed, the focus of much Archaean palaeontology is focussed on whether putative fossils are real or geological artefact. However, evolutionary history is also recorded in the genomes of living organisms and the nature of our long extinct ancestors can be inferred through comparisons to our living relatives. For example, similarties in the genes present in our cells and those of our gut bacteria, reflect aspects of the last universal common ancestor. We have used this parallel fossil record in a molecular palaeobiological approach to elucidate LUCA [1] and the process of eukaryogenesis [2] on which all complex life on Earth is predicated.

The nature of the Last Universal Common Ancestor (LUCA), its age, and impact on the Earth system have been the subject of vigorous debate across diverse disciplines, often based on disparate data. Age estimates for LUCA are usually based on the fossil record, varying with every reinterpretation. The nature of LUCA's metabolism has proven equally contentious, with some attributing all core metabolisms to LUCA, while others reconstruct a simpler life form dependent on geochemistry. Here, we infer that LUCA lived ~4.2 Ga (4.09 - 4.33 Ga) through divergence time analysis of pre-LUCA gene duplicates, calibrated using microbial fossils and isotope records under a new cross-bracing implementation. Phylogenetic reconciliation suggests that LUCA had a genome of at least 2.5Mb (2.49-2.99Mb), encoding around 2500 proteins, comparable to modern prokaryotes. Our results suggest LUCA was a prokaryote-grade anaerobic acetogen that possessed an early immune system. While LUCA is sometimes perceived as living in isolation, we infer LUCA to have been part of an established ecological system. The metabolism of LUCA would have provided a niche for other microbial community members, while hydrogen recycling by atmospheric photochemistry could have supported a modestly productive early ecosystem.

The origin of eukaryotes was a formative, but poorly understood event in the history of life. Current hypotheses of eukaryogenesis differ principally on the timing of mitochondrial endosymbiosis relative to the acquisition of other eukaryote novelties. Discriminating among these has been challenging because there are no living lineages representative of intermediate steps within eukaryogenesis. However, many eukaryotic cell functions are contingent on genes that emerged from duplication events during eukaryogenesis. Consequently, the timescale of these duplications can provide insights into the sequence of steps in the evolutionary assembly of the eukaryotic cell. Here we show, using a relaxed molecular clock, that the process of eukaryogenesis spanned the Mesoarchaean to late Palaeoproterozoic. Within these constraints, we dated the timing of these gene duplications, revealing that host cell was already very complex before mitochondrial endosymbiosis, including an elaborated cytoskeleton, membrane trafficking, endomembrane, phagocytotic machinery, and a nucleus, all between 3.0-2.25 Ga, after which mitochondrial endosymbiosis occurred. As such, our results allow us to reject mitochondrion-early scenarios of eukaryogenesis, instead supporting a complexified-archaean, late-mitochondrion (CALM) pattern for the assembly of eukaryote characteristics. Our inference of a complex archaean host cell is compatible with hypotheses on the adaptive benefits of syntrophy, in oceans that would have remained largely anoxic for more than a billion years.

[1] E. R.R. Moody et al., *Nat. Ecol. Evol.* **2024,** *8,* 1654-1666.
[2] C. J. Kay et al., *Nature* (in press).