



# A New View of the Last Universal Common Ancestor

Aaron D. Goldman<sup>1,2</sup> · Arturo Becerra<sup>3</sup>

Received: 26 July 2024 / Accepted: 26 July 2024  
© The Author(s) 2024

**Keywords** Last universal common ancestor · Time tree analysis · Universal paralogs · Tree reconciliation · Ancient metabolism · Ancient ecology

The last universal common ancestor (LUCA) is one of the earliest stages of evolution that can be studied using standard phylogenetic methods. Recent decades have seen numerous studies aiming to date LUCA's emergence (Betts et al. 2018), reconstruct its genome and proteome (Becerra et al. 2007; Goldman et al. 2013), and predict its metabolism (Weiss et al. 2016; Goldman et al. 2023; Ledford and Meredith 2024), habitat (Weiss et al. 2016), and ecological context (Krupovic et al. 2020). While preparing the special issue on the LUCA for the Journal of Molecular Evolution, a new and innovative study by Moody et al. (2024) was published. We describe this study and situate it within the broader context of early evolutionary history.

The study by Moody et al. presents several analyses on the timing, proteome, genome, environment, and ecology of the LUCA. Dating the LUCA or any ancient ancestor is a complex and uncertain research problem. Molecular clocks based on constant mutation rates are unreliable due to variations over time and across lineages. Proteins used to date the LUCA must also be under strong negative selection, further obscuring the relationship between the number of mutations and geological time. A recent method calibrates molecular clocks using horizontal gene transfers that can be identified on phylogenetic trees and that correspond to a geochemical signature or fossil that can be dated (Wolfe and Fournier 2018). Moody et al. take advantage of similar fossil-based

time constraints (Betts et al. 2018) while also calibrating their molecular clock using universal paralog protein families (Shih and Matzke 2013).

Universal paralogs are protein families that duplicated before the last universal common ancestor (LUCA), resulting in paralogous proteins in the LUCA proteome (Kollman and Doolittle 2000; Zhaxybayeva et al. 2005). In a phylogenetic tree of a universal paralog family, each paralog forms a separate clade with its own LUCA node. By comparing the time estimate of the LUCA node across both paralog subtrees of the same family, the time estimate can be cross-calibrated. Moody et al. applied calibrated molecular clock analysis to a collection of five well-established universal paralog trees and estimate the date of the LUCA to be ~4.2 Ga (4.09–4.32 Ga). This dating is interesting because it leaves very little time for the LUCA to emerge. The moon-forming impact likely occurred around 4.5 Ga (Fu et al. 2023), and the first habitable environments likely formed around 4.3 or 4.4 Ga (Wilde et al. 2001; Miyazaki and Korenaga 2022). Therefore, the 4.2 Ga dating, if correct, suggests that early evolution, from the origin of life to the LUCA, may have occurred within only 100–200 million years.

This finding is particularly surprising given that the LUCA also appears to represent an evolutionary stage in which organisms were already as complex as some modern prokaryotes (Becerra et al. 2007). Numerous studies have attempted to reconstruct the genome or proteome of the LUCA, and most depict it as possessing a DNA genome, a complete translation system, a cell membrane, and a complex metabolism. However, despite this general consensus, the specific predictions of individual studies often diverge, rendering them somewhat unreliable (Crapitto et al. 2022). This inconsistency likely results from a lack of best practices in LUCA proteome reconstruction studies. Specifically, no previous LUCA proteome study has employed both taxonomically broad sampling and paired gene tree-species tree

Handling editor: **David Liberles**.

✉ Aaron D. Goldman  
agoldman@oberlin.edu

✉ Arturo Becerra  
abb@ciencias.unam.mx

<sup>1</sup> Department of Biology, Oberlin College, Oberlin, OH, USA

<sup>2</sup> Blue Marble Space Institute of Science, Seattle, WA, USA

<sup>3</sup> Facultad de Ciencias, UNAM, Mexico City, Mexico

comparisons, which are essential for detecting horizontal gene transfers.

In this context, the Moody et al. study represents a significantly more rigorous approach to LUCA proteome reconstructions. The study used tree reconciliation analysis on 9365 protein families, comparing each protein tree to a species tree to estimate gene duplications, gene losses, and horizontal gene transfers (Szöllösi et al. 2013). This allowed the authors to assign probabilities of being present in the LUCA proteome for each protein family. Using the strictest criteria, 399 protein families were identified as present in the LUCA proteome, which is similar in number to other recent reconstructions, such as 355 protein families from Weiss et al. (2016) and 366 protein families from Crapitto et al. (2022). However, Moody et al. also used the probabilities for all 9365 protein families to estimate that the LUCA proteome likely comprised 2451–2855 proteins. This estimate, though much higher than previous reconstructions, aligns with the idea that LUCA had a complexity similar to modern, free-living prokaryotes, which typically encode thousands of proteins (Sela et al. 2016).

This probable LUCA proteome allowed the authors to describe the LUCA's metabolism, physiology, habitat, and ecology. The metabolic reconstruction includes pathways common across the tree of life, such as glycolysis/gluconeogenesis, the citric acid cycle, and nucleotide biosynthesis. It also features a nearly complete Wood–Ljungdahl Pathway, used by modern archaea and bacteria for acetogenic growth and carbon fixation. The presence of this pathway could indicate an organoheterotrophic acetogenic metabolism, further suggesting that the LUCA would necessarily have been part of an ecosystem with autotrophic organisms that provided it with organic compounds. If, on the other hand, the LUCA had a chemoautotrophic acetogenic metabolism, the authors suggest that its most likely habitats would have been either hydrothermal vents or the ocean surface. However, key enzymes in early CO<sub>2</sub> fixation pathways and the overlap between methanogenic, acetogenic, and autotrophic pathways, such as the CO dehydrogenase/acetyl-CoA synthase, complicate a definitive conclusion about LUCA's metabolism (Becerra et al. 2014). Despite this, the study by Moody et al., with its comprehensive phylogenetic representation of basal lineages in both prokaryotic domains, is a significant step toward uncovering the metabolic capabilities of the LUCA and determining whether it relied on acetogenic growth or carbon fixation.

The reconstructed LUCA proteome includes physiological features such as DNA synthesis, aminoacyl-tRNA biosynthesis, ribosome formation, a cell membrane, and membrane-based ATP synthesis. One of the more surprising results of the reconstructed LUCA physiology is the presence of a CRISPR-Cas system, indicating that an immune system against viral infection had already evolved

by the time of the LUCA. Therefore, Moody et al. depict the LUCA as a complex cellular organism with a sophisticated genetic system, advanced energy metabolism, and viral defense mechanisms. This LUCA likely lived in one of two potential ocean environments, coexisting with other organisms whose descendants either did not survive or have yet to be discovered. This depiction of a physiologically and ecologically complex LUCA provides strong evidence that this stage of evolutionary history was far removed from the origin of life.

This new study of the LUCA marks a significant advance in our understanding. However, although we have a probable list of protein families present in the LUCA, further research is required to determine the specific molecular functions of their ancient ancestors. Understanding these functions would not only support predictions about the LUCA's habitat and ecology but also deepen our understanding of early evolutionary processes. Additionally, a simplified characterization of the LUCA, focusing on traits under strong natural selection, is generally more robust and testable than a detailed catalog of metabolic genes. Viewing the LUCA as a population of cells, and the reconstructed LUCA proteome as representing a core genome of a broader pangenome (Goldman and Kaçar 2023), can lead to valuable insights into early evolution and diversification mechanisms (Estrada et al. 2022). To this end, the disparity described by Moody et al. between the probable LUCA proteome (399 protein families) and the estimated number of encoded proteins in the LUCA genome (2451–2855 proteins) offers an enticing lead toward a broader, more nuanced understanding of the LUCA.

In sum, Moody et al. provide a detailed and reliable depiction of the LUCA. But in doing so, their study raises important new questions that will no doubt be the focus of early evolution research in the near future.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Becerra A, Delaye L, Islas S et al (2007) The very early stages of biological evolution and the nature of the last common ancestor of the three major cell domains. *Annu Rev Ecol Evol Syst* 38:361–379. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095825>

- Becerra A, Rivas M, García-Ferris C et al (2014) A phylogenetic approach to the early evolution of autotrophy: the case of the reverse TCA and the reductive acetyl-CoA pathways. *Int Microbiol* 17(2):91–97. <https://doi.org/10.2436/20.1501.01.211>
- Betts HC, Puttick MN, Clark JW et al (2018) Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat Ecol Evol* 2:1556–1562
- Crapitto AJ, Campbell A, Harris AJ, Goldman AD (2022) A consensus view of the proteome of the last universal common ancestor. *Ecol Evol* 12(6):e8930. <https://doi.org/10.1002/ece3.8930>
- Estrada A, Suárez-Díaz E, Becerra A (2022) Reconstructing the last common ancestor: epistemological and empirical challenges. *Acta Biotheor* 70:15. <https://doi.org/10.1007/s10441-022-09439-1>
- Fu H, Jacobsen SB, Sedaghatpour F (2023) Moon's high-energy giant-impact origin and differentiation timeline inferred from Ca and Mg stable isotopes. *Commun Earth Environ* 4:307. <https://doi.org/10.1038/s43247-023-00974-4>
- Goldman AD, Bernhard TM, Dolzhenko E, Landweber LF (2013) LUCApedia: a database for the study of ancient life. *Nucleic Acids Res* 41:D1079–1082. <https://doi.org/10.1093/nar/gks1217>
- Goldman AD, Kaçar B (2023) Very early evolution from the perspective of microbial ecology. *Environ Microbiol* 25(1):5–10. <https://doi.org/10.1111/1462-2920.16144>
- Goldman AD, Weber JM, LaRowe DE, Barge LM (2023) Electron transport chains as a window into the earliest stages of evolution. *Proc Natl Acad Sci U S A* 120(34):e2210924120. <https://doi.org/10.1073/pnas.2210924120>
- Kollman JM, Doolittle RF (2000) Determining the relative rates of change for prokaryotic and eukaryotic proteins with anciently duplicated paralogs. *J Mol Evol* 51:173–181
- Krupovic M, Dolja VV, Koonin EV (2020) The LUCA and its complex virome. *Nat Rev Microbiol* 18(11):661–670. <https://doi.org/10.1038/s41579-020-0408-x>
- Ledford SM, Meredith LK (2024) Volatile organic compound metabolism on early Earth. *J Mol Evol*. <https://doi.org/10.1007/s00239-024-10184-x>
- Miyazaki Y, Korenaga J (2022) A wet heterogeneous mantle creates a habitable world in the Hadean. *Nature* 603:86–90. <https://doi.org/10.1038/s41586-021-04371-9>
- Moody ERR, Álvarez-Carretero S, Mahendrarajah TA et al (2024) The nature of the last universal common ancestor and its impact on the early Earth system. *Nat Ecol Evol*. <https://doi.org/10.1038/s41559-024-02461-1>
- Sela I, Wolf YI, Koonin EV (2016) Theory of prokaryotic genome evolution. *Proc Natl Acad Sci U S A* 113(41):11399–11407. <https://doi.org/10.1073/pnas.1614083113>
- Shih PM, Matzke NJ (2013) Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *Proc Natl Acad Sci* 110(30):12355–12360. <https://doi.org/10.1073/pnas.1305813110>
- Szöllösi GJ, Rosikiewicz W, Boussau B et al (2013) Efficient exploration of the space of reconciled gene trees. *Syst Biol* 62(6):901–912. <https://doi.org/10.1093/sysbio/syt054>
- Weiss MC, Sousa FL, Mrnjavac N et al (2016) The physiology and habitat of the last universal common ancestor. *Nat Microbiol* 1(9):16116. <https://doi.org/10.1038/nmicrobiol.2016.116>
- Wilde SA, Valley JW, Peck WH, Graham CM (2001) Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* 409:175–178
- Wolfe JM, Fournier GP (2018) Horizontal gene transfer constrains the timing of methanogen evolution. *Nat Ecol Evol* 2(5):897–903
- Zhaxybayeva O, Lapierre P, Gogarten JP (2005) Ancient gene duplications and the root(s) of the tree of life. *Protoplasm* 227:53–64