SCIENCE CHINA
Earth Sciences

•HIGHLIGHT•

June 2025 Vol.68 No.6: 2044–2047
https://doi.org/10.1007/s11430-024-1556-6

Inferring ancestral age and metabolic traits—Promises and challenges in the study of the evolution of ancient life

Wenkai TENG & Chuanlun ZHANG*

Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

Received December 10, 2024; revised March 27, 2025; accepted March 28, 2025; published online April 28, 2025

Citation: Teng W, Zhang C. 2025. Inferring ancestral age and metabolic traits—Promises and challenges in the study of the evolution of ancient life. Science China Earth Sciences, 68(6): 2044–2047, https://doi.org/10.1007/s11430-024-1556-6

All extant cellular organisms derive from a common ancestor, which can be illustrated by the universality of many pivotal metabolic traits, including the utilization of ATP as an energy currency and multiple mechanisms processing the genetic information flow as demonstrated by the Central Dogma of biology (Theobald, 2010; Hug et al., 2016). Therefore, studying the age and metabolic traits of ancient life forms, especially the last universal common ancestor (LUCA), is essential to understanding the origin and evolutionary history of life on Earth (Crapitto et al., 2022). Recently, Moody et al. (2024) declared that LUCA lived around 4.2 billion years ago (Ga) and had a genome of at least 2.5 Mb, which was inferred based on the latest dataset and evolutionary models to enhance accuracy. The LUCA genome was suggested to contain up to 2,600 protein-coding genes which was comparable to an extant prokaryotic genome. Furthermore, they innovatively suggested that LUCA likely harbors an acetogenic metabolism with CO_2 and H_2 as substrates and provides niches for other members in an established ecological system. (Moody et al., 2024).

Characterization of the age and metabolic traits of LUCA has become a bottleneck restricting many important studies in both life and earth sciences. Firstly, relaxed molecular clock methods have been widely employed to date the ancestral nodes in phylogenetic trees, an approach that is highly sensitive to fossil and geochemical data (Ho and Duchêne, 2014). However, Precambrian fossil and geochemical data are very scarce and ambiguous, resulting in large errors on dating the root node of a phylogenetic tree, which usually represents LUCA (Moody et al., 2022). According to the age of the Moon-forming impact and the oldest fossil record of ancient life, the root node of the tree of life could only be constrained with a large time horizon (4.5–3.4 Ga) and LUCA was also estimated to live between 4.5 Ga and 3.4 Ga in previous studies (Sugitani et al., 2015; Betts et al., 2018). To address these challenges, Moody et al. (2024) employed pre-LUCA paralogs to improve the reliability of molecular clocks. Just as modern genomes contain paralogs, certain genes (e.g., ATPase) may have undergone gene duplication prior to the emergence of LUCA and retained as paralogous homologs within LUCA's genome (Gogarten and Taiz, 1992). By utilizing these pre-LUCA duplicated genes, LUCA will be represented by two descendant nodes instead of one root node in the phylogenetic tree (Figure 1a). Cross-bracing of calibration by applying identical fossil constraints to two mirrored nodes from two subtrees, simultaneously, could potentially reduce the unreliability of the molecular clock model caused by the rate heterogeneity of sequence evolution (Gribaldo and Cammarano, 1998; Moody et al., 2024). Moody et al. (2024) employed five pairs of pre-LUCA paralogs to carry out multiple fossil calibrations with both concatenated and partitioned sequence alignments. This approach generated the most precise inference to date for the age of LUCA (95% confidence interval [CI], 4.33–4.18 Ga).

* Corresponding author (email: zhangcl@sustech.edu.cn)

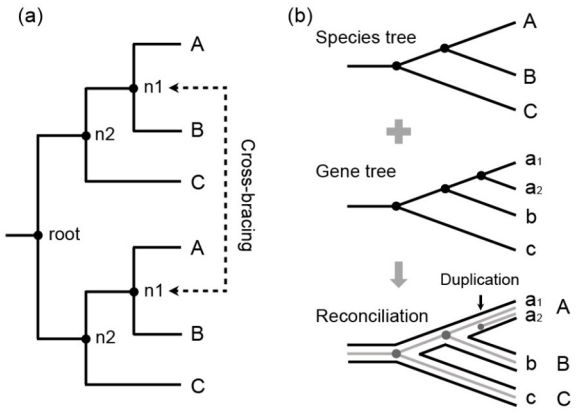

Figure 1 Schematic diagram of timetree inference with cross-bracing and gene tree-species tree reconciliation. (a) For simplicity, we assume 3 species A, B, and C. Phylogenetic reconstruction of a gene that has duplicated before the common ancestor (CA) will generate a tree containing two mirrored subtrees, where two mirrored nodes (n2) instead of the root represent CA. The same fossil calibration can be applied to two mirrored nodes (n1, for example) simultaneously (Moody et al., 2024). (b) Assuming that a_1, a_2, b and c are homologous genes. While b and c are from species B and C, respectively, both a_1 and a_2 are from species A (i.e., a_1 and a_2 are paralogs). Reconciliation of the species tree and gene tree will suggest a gene duplication event in the branch ahead of A (Szöllősi et al., 2013).

Moody et al. (2024) also provided new insights on LUCA's metabolic traits. Early studies reconstructed the genome content of LUCA using gene distribution patterns (Kunin and Ouzounis, 2003). Considering vertical gene descent only, researchers reconstructed an over-simplistic "LUCA genome" based on universally conserved genes. This "LUCA genome" encoded ribosomal proteins and other essential proteins involved in DNA replication, gene transcription, and protein translation, but lacked sufficient enzymes for a more detailed metabolic reconstruction (Harris et al., 2003). Moody et al. (2024) inferred the evolution history of each KEGG Orthology (KO) utilizing a relatively new approach called amalgamated likelihood estimation (ALE). Specifically, ALE reconciled the gene trees of all gene families and a species tree (constructed using concatenated alignments of conserved single-copy genes) to identify events in evolutionary history, including gene duplications, horizontal gene transfers, and gene losses, and to estimate the probability of a gene family presenting in a node in the species tree (Figure 1b).

By analyzing 700 genomes (350 Archaea and 350 Bacteria), Moody et al. (2024) identified 399 KOs with high presence probabilities (PPs, ⩾0.75) in LUCA. These KOs were also found in both modern Archaea and Bacteria. With a lower threshold (PPs⩾0.75), the number of KOs that were likely to be present in LUCA was up to 1,124, and based on this, Moody et al. (2024) further inferred a genome size around 2.75 Mb (95% CI, 2.49–2.99 Mb). In contrast to the over-simplistic LUCA genome in previous studies (Crapitto et al., 2022), Moody et al. (2024) described a genome profile about LUCA that was similar in genome content to modern prokaryotes. The authors posited that LUCA most likely inhabited ocean surfaces or deep-sea environments and was equipped with the Wood-Ljungdahl pathway for acetogenesis, gluconeogenesis/glycolysis for anaerobic growth, and even an early RNA-based immune system. Given a complex genome and metabolic versatility, LUCA was suggested to be a part of the early Earth ecological system regardless of its autotrophic or heterotrophic lifestyle.

Collectively, the methodological framework and findings demonstrated by Moody et al. (2024) regarding the age and metabolic traits of LUCA have provided critical insights for studying the evolutionary history of ancient life on Earth. Contemporary phylogenetic analyses employing gene tree-species tree reconciliation have enabled the inference of ancestral metabolic traits of multiple lineages, including the total bacteria, the closely clustered thermophilic bacteria (CCTB) clade, and archaeal lineages such as Nitrososphaeria, Bathyarchaeia, Heimdallarchaeia, and Marine group II (MGII) archaea (Coleman et al., 2021; Qi et al., 2021; Eme et al., 2023; Leng et al., 2023; Fan et al., 2024; Luo et al., 2024). In combination with molecular dating of ancestral nodes, scientists are progressively establishing a comprehensive framework for understanding the relationship between biospheric evolution and planetary-scale geochemical processes (Hou et al., 2023).

However, several challenges remain in contemporary studies, including the limited number of pre-LUCA paralogs and uncertainties regarding their genetic stability (Zhaxybayeva et al., 2005). In other words, whether these paralogs are all of stable vertical descent (i.e., orthologs) after LUCA requires further investigation. Moreover, gene families with comprehensive functional descriptions (i.e., assigned with KO numbers) constitute only a subset (approximately 55% on average) of all coding genes in modern bacterial and archaeal genomes, with notable lineage-specific variations in annotation proportion (Figure 2). This leads to a major knowledge gap in our understanding of microbial metabolic capabilities and diversity, which also imposes constraints on the estimation of ancestral metabolic traits (Koppel et al., 2018; Crapitto et al., 2022). For example, while Moody et al. (2024) reconstructed a LUCA genome encoding over 2,600 proteins, only 1,124 (with PPs⩾0.5) of them were of known biological functions and merely 399 (with PPs⩾0.75) of them were used to reconstruct the metabolic pathways.

The investigation of microbial genomic "dark matter" and the ancestral genome reconstruction from more prokaryotic lineages (particularly those where major evolutionary events have occurred) may progressively help to reconstruct a more authentic LUCA. Beyond the cross-bracing approach employed by Moody et al. (2024), recent studies have established phylogenetic trees incorporating bacteria, mitochondria, and plastids, thus allowing the molecular

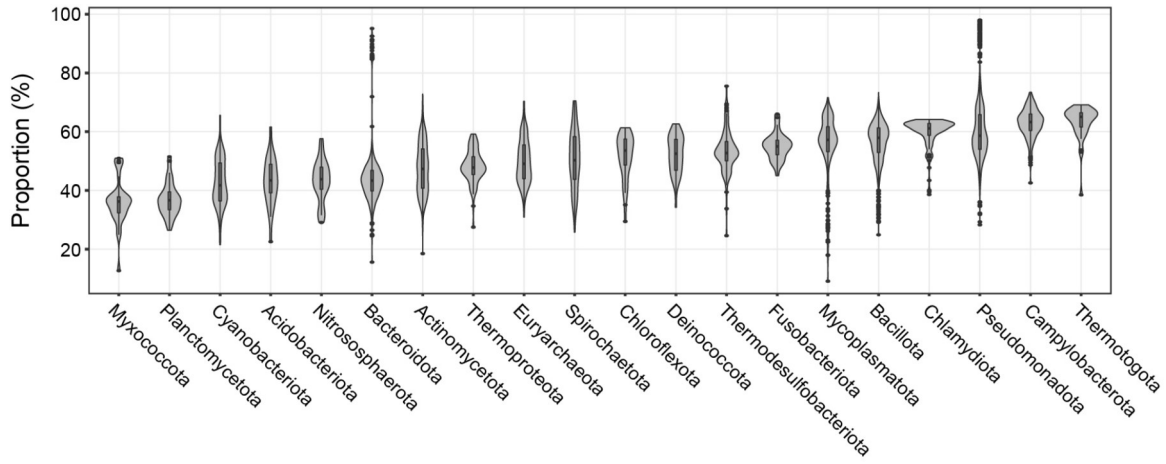

Figure 2 Proportion of genes assigned with KOs to all genes in a genome from different phyla. This result was generated based on the latest KEGG database in 2024 (https://www.genome.jp/kegg/).

dating of prokaryotic nodes by leveraging abundant eukaryotic fossil records (Liao et al., 2024). These efforts potentially improved the reliability of the molecular clock. Notably, encouraging progress regarding the coupling between the evolution of metabolic traits (including ammonia oxidation, anaerobic methane oxidation, and carboxylate oxidation) and major geological events has also been made through molecular clock applications (Ren et al., 2019; Yang et al., 2021; Wang et al., 2022; Shang, 2023). Overall, with the rapid advancement of big-data analytics and artificial intelligence, interdisciplinary collaboration among scientists is expected to yield novel and more accurate standardized methodologies. These efforts will facilitate better explorations of both the temporal origins and metabolic characteristics of prokaryotic ancestors and the co-evolutionary mechanisms between life and geochemical processes during major geological events.

Acknowledgements *Preparation of this article benefited significantly from the comments provided by Prof. Tom A WILLIAMS and Dr. Edmund RR MOODY from the University of Bristol, Prof. Lu FAN from Southern University of Science and Technology, Prof. Yinzhao WANG from Shanghai Jiao Tong University, and Dr. Haitao SHANG from the University of Texas. This work was supported by the National Natural Science Foundation of China (Grant Nos. 92351301, 32393974) and the Science and Technology Innovation Committee of Shenzhen Municipality (Grant No. 20200925173954005).*

Conflict of interest The authors declare that they have no conflict of interest.

References

Betts H C, Puttick M N, Clark J W, Williams T A, Donoghue P C J, Pisani D. 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nat Ecol Evol, 2: 1556–1562

Coleman G A, Davín A A, Mahendrarajah T A, Szánthó L L, Spang A, Hugenholtz P, Szöllősi G J, Williams T A. 2021. A rooted phylogeny resolves early bacterial evolution. Science, 372: eabe0511

Crapitto A J, Campbell A, Harris A J, Goldman A D. 2022. A consensus view of the proteome of the last universal common ancestor. Ecol Evol, 12: e8930

Eme L, Tamarit D, Caceres E F, Stairs C W, De Anda V, Schön M E, Seitz K W, Dombrowski N, Lewis W H, Homa F, Saw J H, Lombard J, Nunoura T, Li W J, Hua Z S, Chen L X, Banfield J F, John E S, Reysenbach A L, Stott M B, Schramm A, Kjeldsen K U, Teske A P, Baker B J, Ettema T J G. 2023. Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. Nature, 618: 992–999

Fan L, Xu B, Chen S, Liu Y, Li F, Xie W, Prabhu A, Zou D, Wan R, Li H, Liu H, Liu Y, Kao S J, Chen J, Zhu Y, Rinke C, Li M, Zhu M, Zhang C, Dupont C. 2024. Gene inversion led to the emergence of brackish archaeal heterotrophs in the aftermath of the Cryogenian Snowball Earth. Proc Natl Acad Sci U S A Nexus, 3: pgae057

Gogarten J P, Taiz L. 1992. Evolution of proton pumping ATPases: Rooting the tree of life. Photosynth Res, 33: 137–146

Gribaldo S, Cammarano P. 1998. The root of the universal tree of life inferred from anciently duplicated genes encoding components of the protein-targeting machinery. J Mol Evol, 47: 508–516

Harris J K, Kelley S T, Spiegelman G B, Pace N R. 2003. The genetic core of the universal ancestor. Genome Res, 13: 407–412

Ho S Y W, Duchêne S. 2014. Molecular-clock methods for estimating evolutionary rates and timescales. Mol Ecol, 23: 5947–5965

Hou J, Wang Y, Zhu P, Yang N, Liang L, Yu T, Niu M, Konhauser K, Woodcroft B J, Wang F. 2023. Taxonomic and carbon metabolic diversification of Bathyarchaeia during its coevolution history with early Earth surface environment. Sci Adv, 9: eadf5069

Hug L A, Baker B J, Anantharaman K, Brown C T, Probst A J, Castelle C J, Butterfield C N, Hernsdorf A W, Amano Y, Ise K, Suzuki Y, Dudek N, Relman D A, Finstad K M, Amundson R, Thomas B C, Banfield J F. 2016. A new view of the tree of life. Nat Microbiol, 1: 16048

Koppel N, Bisanz J E, Pandelia M E, Turnbaugh P J, Balskus E P. 2018. Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. eLife, 7: e33953

Kunin V, Ouzounis C A. 2003. GeneTRACE–reconstruction of gene content of ancestral species. Bioinformatics, 19: 1412–1416

Leng H, Wang Y, Zhao W, Sievert S M, Xiao X. 2023. Identification of a deep-branching thermophilic clade sheds light on early bacterial evolution. Nat Commun, 14: 4354

Liao T, Wang S, Zhang H, Stueken E E, Luo H, Battistuzzi F U. 2024. Dating ammonia-oxidizing bacteria with abundant eukaryotic fossils. Mol Biol Evol, 41: msae096

Luo Z H, Li Q, Xie Y G, Lv A P, Qi Y L, Li M M, Qu Y N, Liu Z T, Li Y

X, Rao Y Z, Jiao J Y, Liu L, Narsing Rao M P, Hedlund B P, Evans P N, Fang Y, Shu W S, Huang L N, Li W J, Hua Z S. 2024. Temperature, pH, and oxygen availability contributed to the functional differentiation of ancient *Nitrososphaeria*. ISME J, 18: wrad031

Moody E R R, Álvarez-Carretero S, Mahendrarajah T A, Clark J W, Betts H C, Dombrowski N, Szánthó L L, Boyle R A, Daines S, Chen X, Lane N, Yang Z, Shields G A, Szöllősi G J, Spang A, Pisani D, Williams T A, Lenton T M, Donoghue P C J. 2024. The nature of the last universal common ancestor and its impact on the early Earth system. Nat Ecol Evol, 8: 1654–1666

Moody E R, Mahendrarajah T A, Dombrowski N, Clark J W, Petitjean C, Offre P, Szöllősi G J, Spang A, Williams T A. 2022. An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. eLife, 11: e66695

Qi Y L, Evans P N, Li Y X, Rao Y Z, Qu Y N, Tan S, Jiao J Y, Chen Y T, Hedlund B P, Shu W S, Hua Z S, Li W J, Lindemann S R. 2021. Comparative Genomics Reveals Thermal Adaptation and a High Metabolic Diversity in "*Candidatus* Bathyarchaeia". mSystems, 6: 10.1128/msystems.00252-21

Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, Swan B K, Fonseca M M, Posada D, Stepanauskas R, Hollibaugh J T, Foster P G, Woyke T, Luo H. 2019. Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. ISME J, 13: 2150–2161

Shang H. 2023. Dichotomous effects of oxidative metabolisms: A theoretical perspective on the dolomite problem. Glob Planet Change, 222: 104041

Sugitani K, Mimura K, Takeuchi M, Lepot K, Ito S, Javaux E J. 2015. Early evolution of large micro-organisms with cytological complexity revealed by microanalyses of 3.4 Ga organic-walled microfossils. Geobiology, 13: 507–521

Szöllősi G J, Rosikiewicz W, Boussau B, Tannier E, Daubin V. 2013. Efficient exploration of the space of reconciled gene trees. Systatic Biol, 62: 901–912

Theobald D L. 2010. A formal test of the theory of universal common ancestry. Nature, 465: 219–222

Wang Y, Xie R, Hou J, Lv Z, Li L, Hu Y, Huang H, Wang F. 2022. The late Archaean to early Proterozoic origin and evolution of anaerobic methane-oxidizing archaea. mLife, 1: 96–100

Yang Y, Zhang C, Lenton T M, Yan X, Zhu M, Zhou M, Tao J, Phelps T J, Cao Z, Ursula Battistuzzi F. 2021. The evolution pathway of ammonia-oxidizing archaea shaped by major geological events. Mol Biol Evol, 38: 3637–3648

Zhaxybayeva O, Lapierre P, Gogarten J P. 2005. Ancient gene duplications and the root(s) of the tree of life. Protoplasma, 227: 53–64

(Editorial handling: Tuo SHI)