

LUCA: The Ancestor We Reconstructed

QTC-023 · From Quarks to Consciousness

The Universe, Life, and How We Found Out

Author: Juan Segurola

Version: v1.0 EN · 2 September 2026

Epistemic profile: Current scientific interpretation

License: All Rights Reserved

Article Abstract

LUCA is not a fossil, the first life or one solitary organism. It is the last ancestral population shared by every surviving cellular lineage, reconstructed from molecular descendants. Ribosomal RNA first exposed the deep bacterial–archaeal division; ancient gene duplications helped root it. Comparative genomics then revealed a small universal core, heavily weighted toward translation, while gene loss and horizontal transfer complicated every larger portrait. Transfer can also become evidence by constraining relative order. Recent analyses combine reconciled gene trees, probabilistic ancestral genomes, microbial fossils, isotope records and relaxed molecular clocks. One 2024 reconstruction places LUCA near 4.2 billion years ago and infers a prokaryote-grade, ecological organism with roughly 2,600 proteins. Those details remain model-dependent. Common ancestry is robust; LUCA's exact genome, membrane, metabolism, habitat and age are layered interpretations.

In 60 Seconds

LUCA means last universal common ancestor. “Last” matters: earlier life may have existed, exchanged genes or vanished. “Common ancestor” means a genealogical population inferred from descendants, not a fossilized individual. The reconstruction starts with shared molecular machinery. Ribosomal RNA divided cellular life deeply into Bacteria and Archaea; duplicated genes older than that split supplied root directions. Universal genes, especially translation components, reveal inherited systems, but a strict intersection misses genes later lost. Simply accepting every cross-domain gene includes later transfers. Modern methods reconcile thousands of gene histories against species frameworks and model gain, loss and transfer. Horizontal transfer is not only noise: its donor and recipient can impose relative chronology. Absolute dates require molecular-rate models plus geological calibrations. In 2024, cross-braced pre-LUCA duplications yielded an age near 4.2 Ga and a probabilistic genome of at least 2.5 Mb, around 2,600 proteins. It also suggested acetogenesis, ecology and early immunity. These are current interpretations, not observations. The membrane remained unresolved, and alternative methods recover different ancestral content.

Why This Question Matters

LUCA is often illustrated as though scientists had found its cell and sequenced its genome. In reality, each feature comes from a different inference with a different failure mode. Common ancestry can be strong while a metabolic pathway remains debatable; a clock can be mathematically precise while its calibrations remain interpretive. Following the reconstruction step by step shows how genomes become historical instruments. It also prevents one attractive drawing from hiding the remarkable

central achievement: extracting a constrained ancestral population from billions of years of mutation, loss, transfer and extinction.

1. An Ancestor Defined by Descendants

The Question

Nothing called LUCA was excavated. The name identifies the most recent ancestral population from which all surviving cellular lineages descend. Its existence is inferred backwards from relationships in living molecules, and every physical trait requires an additional argument.

A common ancestor is defined by descendants. Trace any pair of living lineages backward and their branches eventually meet; expand the sample to all cellular life and the last shared meeting point is LUCA. Douglas Theobald compared models of common and separate ancestry for conserved proteins and found overwhelming support for universal common ancestry in the sampled organisms.[6] The test concerns lineages that survive into the dataset. It does not count every earlier attempt at life.

The word last prevents a common mistake. LUCA need not be the first replicator, first protocell or first cell. Earlier populations could have invented parts of translation, exchanged them and left no independent descendants. Carl Woese described a communal progenote phase in which genetic exchange and immature genotype–phenotype coupling preceded modern cellular lineages.[7,8] That interpretation is not required for common ancestry, but it explains why the last shared ancestor can already contain a long prehistory.

Nor must LUCA have been one genetically uniform cell. A species ancestor is usually a population containing variation, births, deaths and exchange. The ancestral population could have occupied several connected habitats while retaining enough genealogical continuity to seed later bacterial and archaeal lines. Calling it a point on a diagram compresses that population into a node. The compression is useful; mistaking the node for an individual is not.

Shared traits are clues, not automatic possessions. If bacteria and archaea carry homologous ribosomal proteins, the simplest explanation may be inheritance from their common ancestor. A gene can also disappear from many descendants or cross between domains later. Presence on both sides is therefore screened through gene trees, duplication, loss and transfer models.[9,10,15] LUCA is not obtained by listing everything found in two modern cells and taking the overlap.

The reconstruction has layers. Sequence similarity supports homology; aligned homologues support gene trees; gene trees are reconciled with a cellular framework; ancestral-state models assign probabilities to genes and traits; molecular clocks add rates and calibrations to estimate time.[14,18,23] Each layer can preserve uncertainty from the one below. A painted cell at the end should never look more directly observed than the chain that produced it.

Where That Leaves Us

LUCA begins as a relationship, not an organismal portrait. Universal common ancestry identifies a shared ancestral population while leaving its body, habitat and date to further inference. The first decisive map of that relationship came from a molecule no palaeontologist excavated: ribosomal RNA, whose conserved sequence exposed a division deeper than familiar microbial appearances suggested.

2. Ribosomal RNA Revealed the Deep Split

Microbes were once grouped largely by shape, staining, metabolism and whether they possessed a nucleus. Those properties remain useful for identification, but distant relatives can converge on similar forms and close relatives can change metabolism. Carl Woese chose ribosomal RNA because every cellular organism translates proteins, ribosomes contain conserved and variable regions, and sequence differences can accumulate as historical records across enormous spans of time.

In 1977 Woese and George Fox compared ribosomal RNA sequence characteristics and found that methanogens were not an odd branch of ordinary bacteria. They represented a deep line of descent, which they called archaeobacteria.[1] The conclusion reorganized biology because a difference hidden inside ribosomes proved more ancient than many visible traits. Limited organisms and partial oligonucleotide catalogues made the original analysis modest by genomic standards, but later sequences confirmed the central division.

Woese, Otto Kandler and Mark Wheelis formalized Bacteria, Archaea and Eucarya as three domains in 1990.[2] The diagram has since changed. The eocyte proposal and later Asgard archaeal genomes support a close or nested relationship between eukaryotes and Archaea rather than three equally independent primary stems.[20,27,28] That revision changes the placement of eukaryotes. It does not reunite Archaea with Bacteria or make the original molecular discovery disappear.

A sequence tree is not produced by counting mismatches alone. Researchers align homologous positions and choose models for substitutions, rate variation and compositional differences. Fast-evolving sites can erase old signal; unrelated lineages can converge in base composition; misalignment can invent similarity. Modern phylogenomics uses many markers and site-heterogeneous models to reduce these problems.[19–21] More genes increase information only when their histories and errors are not blindly treated as identical.

Ribosomal RNA and conserved translation genes establish a powerful unrooted framework. An unrooted tree says which branches connect but not where time enters. Rotate it on the page and its relationships remain the same. To say that one lineage diverged before another—or to map traits into LUCA—requires a root. No living cellular outgroup stands outside all living cells, so the universal root had to be manufactured from events older than the split itself.

Where That Leaves Us

Ribosomal comparisons established deep relatedness and the bacterial–archaeal divide, but direction was missing. The solution was conceptually elegant: find genes that duplicated before LUCA, then use each copy as an outgroup for its partner. That manoeuvre turns an event older than the ancestral population into an arrow pointing through the universal tree.

3. The Root Comes from Older Duplications

The Question

To root a family tree, biologists normally compare a relative known to lie outside it. Universal cellular ancestry contains every known cell, so its outgroup cannot be another cell. An ancient duplication can create the comparison internally.

Suppose one ancestral gene duplicated into copies A and B before the bacterial–archaeal divergence. Later, both domains inherited descendants of both copies. The A sequences form one family and B sequences another. Because the A–B split is older than the species split, family B can root the tree inside family A, and vice versa. The logic is strong only if the duplication is truly ancient and the copies were not later redistributed in misleading ways.

Naoyuki Iwabe and colleagues applied this strategy to duplicated elongation factors and ATPase-related proteins in 1989. Their paralogue trees placed the root between Bacteria and a branch uniting Archaea with eukaryotes.[3] The result supplied direction that ribosomal similarity alone could not. It also supported an archaeal–eukaryotic relationship before the genomic era. Alignment across ancient paralogues and sparse early sampling remained substantial limitations.

Johann Peter Gogarten and colleagues independently analysed the two related ATPase families, one represented by vacuolar-type enzymes and the other by F-type enzymes.[4] They recovered the same broad root. James Brown and W. Ford Doolittle later used ancient aminoacyl-tRNA synthetase duplications as another window.[5] Agreement among independent duplications matters because every single gene family can carry its own transfer, rate and functional-divergence problems.

Paralogue rooting is not a mechanical truth stamp. A duplication placed on the wrong side of LUCA reverses the logic. Hidden gene loss can leave misleading samples; horizontal transfer can move one copy across domains; long branches can attract. Gregory Fournier and Gogarten revisited ribosomal rooting with explicit models and different ancient signals.[23] Modern analyses test several gene families, taxon sets and substitution models rather than declaring one ancient enzyme decisive.

Root location and common ancestry must stay separate. Theobald's comparison supports one ancestry for the conserved proteins he tested even if researchers debate the exact deepest direction.[6] Rooting affects the order of early divergences and which state a model assigns to LUCA. It does not turn shared translation machinery into unrelated inventions. Uncertainty in an arrow on the tree is not uncertainty about whether a tree-like inheritance signal exists at all.

Where That Leaves Us

Ancient duplications orient the deepest split, yet they reconstruct no genome by themselves. To infer LUCA's contents, researchers ask which gene families cross the split and how they arrived there. The strict answer—a gene present everywhere—is surprisingly small. The realistic answer must recover genes lost by descendants without mistaking later transfers for ancestral inheritance.

4. The Universal Core Is Small and Biased

Take the intersection of genes in every sequenced cell and the survivors are concentrated in translation: ribosomal proteins, RNA-processing factors and components that connect codons to amino acids. J. Kirk Harris and colleagues found a small genetic core for the universal ancestor, dominated by information-processing functions.[15] The result identifies extraordinarily persistent machinery. It cannot be read as LUCA's complete genome, because universal presence is a harsher requirement than ancestral presence.

Gene loss makes the intersection shrink as sampling grows. A parasite or symbiont can discard biosynthetic genes because its host supplies products. Robert Charlebois and Doolittle called the

statistical response “rescuing the core from extinction”: allow plausible losses rather than requiring every modern genome to retain every ancient gene.[17] The rescue introduces assumptions. Too little loss makes LUCA tiny; too much makes it an implausible owner of almost everything.

Horizontal transfer pushes in the opposite direction. A metabolic gene present in some bacteria and some archaea may be ancient, or it may have crossed the domain boundary later. Boris Mirkin and colleagues formalized gain–loss scenarios for genome histories, while later reconciliation methods added duplications and transfers.[18] Each event has a cost or probability. An ancestral gene list is therefore a model-selected history, not a visible layer common to all genomes.

Minimal genes and ancestral genes answer different questions. Arcady Mushegian and Eugene Koonin compared complete bacterial genomes to derive a candidate minimal set for cellular life. [16] A modern cell can require a recently evolved repair or membrane protein; an ancestral gene can later become dispensable after pathway replacement. Rich media and host dependence also change essentiality. A minimal descendant does not become LUCA when researchers remove enough genes.

The strict core is biased toward systems that resist replacement and transfer. Translation is conserved partly because its many interacting components make wholesale substitution difficult. Metabolic pathways are more modular and mobile. A core-only reconstruction can therefore depict LUCA as rich in information processing but strangely unable to feed itself.[13,14] Transfer-aware and probabilistic methods restore metabolism, but they purchase completeness with additional assumptions.

Where That Leaves Us

The universal core is genuine evidence and an incomplete census. Loss hides ancestral genes; transfer imitates them; functional persistence favours translation over metabolism. The complication does not end reconstruction. It changes the unit of evidence from a shared gene to a gene history—and turns the apparent chaos of horizontal exchange into a forest that can sometimes constrain the tree.

5. Genes Make a Forest Around the Tree

The Question

A gene that jumps between lineages breaks a simple vertical genealogy for that gene. But a directed jump also states that donor and recipient coexisted in a particular order. Reconciled carefully, the violation becomes a historical constraint.

Prokaryotic genomes are mosaics. Genes for resistance, metabolism and ecological adaptation cross lineages by transformation, viruses and mobile elements. Doolittle argued in 1999 that no single tree can represent the histories of all genes as though genomes were sealed lineages.[9] The claim was sometimes inflated into “there is no tree.” A forest of discordant gene trees can coexist with a statistical backbone in more vertically constrained systems.

Species-tree methods now model the discord. Sophie Abby and colleagues compared candidate cellular trees by reconciling lateral transfers and found that transfer patterns could support a tree rather than merely corrupt it.[10] A proposed transfer must connect branches that exist at the

same time. A topology demanding many impossible or unnecessarily complex events scores worse. The method does not make every transfer certain; it uses their collective consistency.

Transfers also contain relative chronology. If lineage A donated a gene to lineage B, the donor branch existed before or during the recipient event. Adrián Davín and colleagues used many such constraints to date the order of branches in the tree of life.[11] John Wolfe and Gregory Fournier applied related logic to methanogenesis, using transfers to constrain when methane-producing lineages existed.[12] Relative order precedes an absolute calendar.

The early tree was probably less rigid than modern textbook branches. Woese proposed a communal phase with extensive innovation sharing, and Kalin Vetsigian, Woese and Nigel Goldenfeld modelled exchange-driven convergence of the genetic code.[7,29] These are interpretations and models, not a recovered traffic log. They explain how collective evolution can stabilize shared systems before lineages become more vertically coherent.

The best current picture is neither one immaculate tree nor an undifferentiated web. Translation genes, gene-content patterns, ancient duplications and transfer reconciliation preserve different amounts of vertical signal.[10,19–21] A tree can summarize cellular descent while edges crossing it record acquired genes. LUCA then represents an ancestral population within a reticulated history. Its reconstructed genome depends on deciding which edges descend through the node and which cross later.

Where That Leaves Us

Horizontal transfer makes early genomes mosaic, but it also supplies order and tests proposed trees. Once gene histories are reconciled, researchers can ask a more ambitious question: what chemistry did the ancestral population perform? The answer cannot be read from universal presence alone. It is reconstructed pathway by pathway, with modern ecology informing—but never directly remembering—the past.

6. Metabolism Is Reconstructed, Not Remembered

A metabolic pathway is more than one homologous gene. Its enzymes must have compatible functions, substrates, cofactors and cellular context. Modern descendants can replace steps, reverse direction or recruit a protein to a new task. Reconstructors therefore build gene families, infer trees, reconcile them with lineages and evaluate whether enough ancestral components form a coherent pathway. The output is a probability-weighted physiology, not a metabolic fossil embedded intact in every cell.

Madeline Weiss and colleagues searched 6.1 million prokaryotic protein-coding genes grouped into 286,514 clusters. Only 355 families—about 0.1 percent—met their phylogenetic criteria for LUCA. [13] From their functions and cofactors they inferred an anaerobic, hydrogen-dependent, carbon-fixing, nitrogen-fixing and thermophilic ancestor rich in iron–sulfur and transition-metal chemistry. The selectivity made the list defensible; modern ecological associations still helped turn genes into habitat.

The Wood–Ljungdahl pathway became central to that portrait. Modern acetogens and methanogens use related routes to fix carbon dioxide with hydrogen through metal-rich enzymes. Its energetic economy fits hydrothermal origin models.[13,25] But a present distribution can reflect

vertical inheritance, loss and transfer, and a pathway assembled from inferred pieces may not have operated in exactly the modern direction. Ancient plausibility is stronger than exclusive historical proof.

Membranes expose a persistent blank. Bacteria and Archaea use broadly different lipid stereochemistry, chains and linkages. The 2024 reconstruction recovered evidence for core cellular apparatus and enzymes associated with both lipid types but did not conclusively identify LUCA's membrane composition.[14] Competing scenarios invoke mixed lipids, later replacement or early divergence.[26] A richly reconstructed metabolism can therefore sit inside an unresolved boundary.

Methods disagree for principled reasons. A strict core minimizes false ancestral assignments but loses genes removed by descendants. A presence–absence model restores probable losses but may push gains too deep. Gene-tree reconciliation can distinguish transfer, duplication and loss, yet ancient trees are noisy.[13–15,18] The right response is not to choose the most detailed portrait. It is to label which traits recur across methods and which appear only under one reconstruction.

Where That Leaves Us

Ancestral metabolism emerges from reconciled histories and functional interpretation, so confidence differs trait by trait. Translation is harder to dislodge than thermophily; cellular machinery is firmer than one membrane recipe. Physiology still lacks a date. To place LUCA on a calendar, sequence divergence must be converted into time using rates that vary and rocks that calibrate without naming LUCA directly.

7. Molecular Clocks Need Geological Anchors

The Question

No sample of LUCA entered a mass spectrometer. A molecular clock estimates its age by combining sequence change, a model of variable evolutionary rates and dated constraints from fossils, isotopes or geological events elsewhere on the tree.

The simplest molecular clock imagines substitutions accumulating at a stable rate. Over billions of years, genes and lineages accelerate, slow, saturate and change constraints. Relaxed clocks allow rates to vary under specified distributions. They still need calibration: a fossil assigned to a branch, an isotope signal tied to a metabolism or a geological event that bounds a lineage. Mathematical precision inside the model cannot make an uncertain calibration direct.

Microbial calibrations are especially difficult. A stromatolite can establish life without identifying a crown group; a methane isotope can indicate a process without naming the organism; a cell shape rarely carries enough diagnostic anatomy. Holly Betts and colleagues integrated genomic and fossil evidence in a relaxed-clock account of early life, illustrating both the reach and dependence of the method.[22] Move a calibration or its phylogenetic placement and downstream node ages move.

Ancient duplications again create leverage. The same pre-LUCA duplication produces matching descendant divergences in two paralogue trees. Edmund Moody and colleagues cross-braced those mirrored nodes so that the analysis assigned them shared time distributions.[14] They used five pre-LUCA paralogues, microbial fossil and isotope calibrations, and alternative relaxed-clock

models. Cross-bracing adds information if the duplicated genes and mirrored species divergences are correctly identified.

Their principal analysis placed LUCA near 4.2 billion years ago, with a reported 4.09–4.33 Ga interval.[14] That is far earlier than the oldest widely accepted fossil communities and leaves a short interval after Earth's formation. It is not a rock dated to LUCA's birthday. The interval is conditional on gene alignments, root and duplication histories, rate priors, calibrations and the clock model; alternative structural uncertainty can exceed the printed posterior band.

Transfers add relative constraints. Davín and colleagues used donor–recipient order to date the tree, and Wolfe and Fournier applied transfer logic to methanogen evolution.[11,12] A clock can combine such before–after information with absolute calibrations. Hidden extinct donors or mistaken direction can mislead it. A responsible LUCA date therefore travels with its interval, calibration set, root, clock family and sensitivity tests rather than as a free-standing number.

Where That Leaves Us

Clocks do not discover a date hidden inside sequences; they negotiate among molecular change, geological anchors and explicit rate models. The 4.2 Ga estimate is a powerful current result precisely because its machinery can be inspected. The same analysis produced the most detailed recent portrait of LUCA, whose resolution now demands equally careful separation between posterior probability and observed fact.

8. A Detailed Portrait Is Still a Model

Moody and colleagues reconstructed a LUCA genome of at least 2.5 megabases, with a reported 2.49–2.99 Mb interval and around 2,600 encoded proteins.[14] That resembles the scale of many modern prokaryotes and is far larger than a strict universal core. The difference follows from method: probabilistic reconciliation allows genes to have been ancestral and later lost or transferred. Detail is gained by modelling histories, not by uncovering more universal genes.

The reconstructed organism was anaerobic and acetogenic, using the Wood–Ljungdahl pathway, and was embedded in an ecosystem whose by-products supported other organisms.[14] It was not a lonely first cell. The analysis also mapped components of an early CRISPR–Cas-like immune system. Cas1 and Cas2 were not reconstructed, so the proposed system differs from a complete modern adaptive CRISPR apparatus. Functional continuity across four billion years remains an inference.

Some features were firmer than others. The analysis supported translation, membrane-bound energy machinery, protein targeting and prokaryote-grade cellular organization. It did not conclusively determine membrane composition.[14,26] Individual inferred gene families carried different posterior probabilities. Compressing them into one illustration makes ribosomes, acetogenesis, immunity and a phospholipid outline look equally certain when they are not.

A complex LUCA does not make life's origin complex by definition. It means substantial evolution preceded the last ancestor shared by surviving lineages. Earlier populations may have been simpler, more communal and more dependent on geochemistry.[7,24,30,31] Patrick Forterre's virus-driven domain hypothesis is one example of a specific, non-consensus route.[30] Competing

proposals belong in the space before and around LUCA, not inside the established common-ancestry result.

The reconstruction succeeds when read in layers. Universal common ancestry is robust. The bacterial–archaeal framework and a translation-rich ancestor are strongly supported. Genome size, exact age and broad cellular complexity are current model-based estimates. Particular pathways, habitat, temperature, immunity and membrane receive different support.[6,13,14,24,31] LUCA is not a photograph. It is an auditable synthesis whose disagreements identify where future genomes, models and ancestral biochemistry can change the picture.

Where That Leaves Us

LUCA is one of science's most ambitious reconstructions: a population inferred from molecular descendants after billions of years of loss and exchange. Its strongest feature is not a vivid face but a visible inference chain. From that ancestral population, the next article follows the dominant reality of life's history—the microbial planet built by bacterial and archaeal descendants.

What We Know, and How

- Conserved molecular data strongly support universal common ancestry among extant cellular lineages.[1,2,6]
- Ancient gene duplications can root the deepest cellular split without an external cellular outgroup.[3–5]
- The strict universal gene core is small and strongly enriched in translation and information-processing functions.[15–17]
- Horizontal transfer creates discordant gene histories but can also constrain species trees and relative chronology.[9–12]
- Recent probabilistic reconstruction supports a prokaryote-grade, ecological LUCA while leaving membrane composition unresolved.[14,26]

What We Still Don't Know

- The genetic diversity, population structure and amount of exchange within the ancestral LUCA population.
- Which inferred genes were vertically inherited through LUCA rather than acquired by deep horizontal transfer.
- LUCA's exact membrane composition, temperature range, habitat and dependence on geochemistry.
- How much uncertainty lies outside reported posterior intervals because alternative roots, calibrations and models were not averaged.
- The duration and organization of cellular evolution between first cells and LUCA.

What Could Change Our Minds

- Independent ancestral reconstructions that converge on the same gene families and functions under different transfer–loss models.
- Experimental resurrection of ancient proteins that confirms or overturns inferred substrates, temperatures and pathway roles.

- New deep-branching genomes that alter bacterial or archaeal roots and the distribution of conserved markers.
- Better microbial calibrations with secure phylogenetic placement and model-averaged molecular-clock analyses.
- A membrane-biosynthesis history that resolves whether LUCA used bacterial, archaeal, mixed or replaced lipids.

If You Remember Only One Thing

LUCA is a reconstructed ancestral population, not first life or a fossil. Common ancestry is robust; its genome, metabolism, habitat and approximately 4.2 Ga age remain layered model-based interpretations.

Related Articles

- QTC-022 — The First Cells We Never Saw
- QTC-024 — The Microbial Planet
- QTC-027 — From Information to Inheritance: Tree, Web and Viruses

References

- [1] Carl R. Woese; George E. Fox. Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms. *Proceedings of the National Academy of Sciences* 74, 5088–5090 (1977). DOI 10.1073/pnas.74.11.5088. <https://doi.org/10.1073/pnas.74.11.5088>
- [2] C R Woese; O Kandler; M L Wheelis. Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* 87, 4576–4579 (1990). DOI 10.1073/pnas.87.12.4576. <https://doi.org/10.1073/pnas.87.12.4576>
- [3] N Iwabe; K Kuma; M Hasegawa; S Osawa; T Miyata. Evolutionary Relationship of Archaeobacteria, Eubacteria, and Eukaryotes Inferred from Phylogenetic Trees of Duplicated Genes. *Proceedings of the National Academy of Sciences* 86, 9355–9359 (1989). DOI 10.1073/pnas.86.23.9355. <https://doi.org/10.1073/pnas.86.23.9355>
- [4] J P Gogarten; H Kibak; P Dittrich; L Taiz; E J Bowman; B J Bowman; M F Manolson; R J Poole; T Date; T Oshima; J Konishi; K Denda; M Yoshida. Evolution of the Vacuolar H⁺-ATPase: Implications for the Origin of Eukaryotes. *Proceedings of the National Academy of Sciences* 86, 6661–6665 (1989). DOI 10.1073/pnas.86.17.6661. <https://doi.org/10.1073/pnas.86.17.6661>
- [5] J R Brown; W F Doolittle. Root of the Universal Tree of Life Based on Ancient Aminoacyl-tRNA Synthetase Gene Duplications. *Proceedings of the National Academy of Sciences* 92, 2441–2445 (1995). DOI 10.1073/pnas.92.7.2441. <https://doi.org/10.1073/pnas.92.7.2441>
- [6] Douglas L. Theobald. A Formal Test of the Theory of Universal Common Ancestry. *Nature* 465, 219–222 (2010). DOI 10.1038/nature09014. <https://doi.org/10.1038/nature09014>
- [7] Carl Woese. The Universal Ancestor. *Proceedings of the National Academy of Sciences* 95, 6854–6859 (1998). DOI 10.1073/pnas.95.12.6854. <https://doi.org/10.1073/pnas.95.12.6854>
- [8] Carl R. Woese. On the Evolution of Cells. *Proceedings of the National Academy of Sciences* 99, 8742–8747 (2002). DOI 10.1073/pnas.132266999. <https://doi.org/10.1073/pnas.132266999>

- [9] W. Ford Doolittle. Phylogenetic Classification and the Universal Tree. *Science* 284, 2124–2128 (1999). DOI 10.1126/science.284.5423.2124. <https://doi.org/10.1126/science.284.5423.2124>
- [10] Sophie S. Abby; Eric Tannier; Manolo Gouy; Vincent Daubin. Lateral Gene Transfer as a Support for the Tree of Life. *Proceedings of the National Academy of Sciences* 109, 4962–4967 (2012). DOI 10.1073/pnas.1116871109. <https://doi.org/10.1073/pnas.1116871109>
- [11] Adrián A. Davín; Eric Tannier; Tom A. Williams; Bastien Boussau; Vincent Daubin; Gergely J. Szöllősi. Gene Transfers Can Date the Tree of Life. *Nature Ecology & Evolution* 2, 904–909 (2018). DOI 10.1038/s41559-018-0525-3. <https://doi.org/10.1038/s41559-018-0525-3>
- [12] Joanna M. Wolfe; Gregory P. Fournier. Horizontal Gene Transfer Constrains the Timing of Methanogen Evolution. *Nature Ecology & Evolution* 2, 897–903 (2018). DOI 10.1038/s41559-018-0513-7. <https://doi.org/10.1038/s41559-018-0513-7>
- [13] Madeline C. Weiss; Filipa L. Sousa; Natalia Mrnjavac; Sinje Neukirchen; Mayo Roettger; Shijulal Nelson-Sathi; William F. Martin. The Physiology and Habitat of the Last Universal Common Ancestor. *Nature Microbiology* 1, 16116 (2016). DOI 10.1038/nmicrobiol.2016.116. <https://doi.org/10.1038/nmicrobiol.2016.116>
- [14] Edmund R. R. Moody; Sandra Álvarez-Carretero; Tara A. Mahendrarajah; James W. Clark; Holly C. Betts; Nina Dombrowski; Lénárd L. Szánthó; Richard A. Boyle; Stuart Daines; Xi Chen; Nick Lane; Ziheng Yang; Graham A. Shields; Gergely J. Szöllősi; Anja Spang; Davide Pisani; Tom A. Williams; Timothy M. Lenton; Philip C. J. Donoghue. The Nature of the Last Universal Common Ancestor and Its Impact on the Early Earth System. *Nature Ecology & Evolution* 8, 1654–1666 (2024). DOI 10.1038/s41559-024-02461-1. <https://doi.org/10.1038/s41559-024-02461-1>
- [15] J. Kirk Harris; Scott T. Kelley; George B. Spiegelman; Norman R. Pace. The Genetic Core of the Universal Ancestor. *Genome Research* 13, 407–412 (2003). DOI 10.1101/gr.652803. <https://doi.org/10.1101/gr.652803>
- [16] A R Mushegian; E V Koonin. A Minimal Gene Set for Cellular Life Derived by Comparison of Complete Bacterial Genomes. *Proceedings of the National Academy of Sciences* 93, 10268–10273 (1996). DOI 10.1073/pnas.93.19.10268. <https://doi.org/10.1073/pnas.93.19.10268>
- [17] Robert L. Charlebois; W. Ford Doolittle. Computing Prokaryotic Gene Ubiquity: Rescuing the Core from Extinction. *Genome Research* 14, 2469–2477 (2004). DOI 10.1101/gr.3024704. <https://doi.org/10.1101/gr.3024704>
- [18] Boris G Mirkin; Trevor I Fenner; Michael Y Galperin; Eugene V Koonin. Algorithms for Computing Parsimonious Evolutionary Scenarios for Genome Evolution, the Last Universal Common Ancestor and Dominance of Horizontal Gene Transfer in the Evolution of Prokaryotes. *BMC Evolutionary Biology* 3, 2 (2003). DOI 10.1186/1471-2148-3-2. <https://doi.org/10.1186/1471-2148-3-2>
- [19] Gareth A. Coleman; Adrián A. Davín; Tara A. Mahendrarajah; Lénárd L. Szánthó; Anja Spang; Philip Hugenholtz; Gergely J. Szöllősi; Tom A. Williams. A Rooted Phylogeny Resolves Early Bacterial Evolution. *Science* 372, eabe0511 (2021). DOI 10.1126/science.abe0511. <https://doi.org/10.1126/science.abe0511>

- [20] Tom A. Williams; Gergely J. Szöllősi; Anja Spang; Peter G. Foster; Sarah E. Heaps; Bastien Boussau; Thijs J. G. Ettema; T. Martin Embley. Integrative Modeling of Gene and Genome Evolution Roots the Archaeal Tree of Life. *Proceedings of the National Academy of Sciences* 114, E4602–E4611 (2017). DOI 10.1073/pnas.1618463114. <https://doi.org/10.1073/pnas.1618463114>
- [21] Laura A. Hug; Brett J. Baker; Karthik Anantharaman; Christopher T. Brown; Alexander J. Probst; Cindy J. Castelle; Cristina N. Butterfield; Alex W. Hernsdorf; Yuki Amano; Kotaro Ise; Yohey Suzuki; Natasha Dudek; David A. Relman; Kari M. Finstad; Ronald Amundson; Brian C. Thomas; Jillian F. Banfield. A New View of the Tree of Life. *Nature Microbiology* 1, 16048 (2016). DOI 10.1038/nmicrobiol.2016.48. <https://doi.org/10.1038/nmicrobiol.2016.48>
- [22] Holly C. Betts; Mark N. Puttick; James W. Clark; Tom A. Williams; Philip C. J. Donoghue; Davide Pisani. Integrated Genomic and Fossil Evidence Illuminates Life's Early Evolution and Eukaryote Origin. *Nature Ecology & Evolution* 2, 1556–1562 (2018). DOI 10.1038/s41559-018-0644-x. <https://doi.org/10.1038/s41559-018-0644-x>
- [23] G. P. Fournier; J. P. Gogarten. Rooting the Ribosomal Tree of Life. *Molecular Biology and Evolution* 27, 1792–1801 (2010). DOI 10.1093/molbev/msq057. <https://doi.org/10.1093/molbev/msq057>
- [24] David Penny; Anthony Poole. The Nature of the Last Universal Common Ancestor. *Current Opinion in Genetics & Development* 9, 672–677 (1999). DOI 10.1016/S0959-437X(99)00020-9. [https://doi.org/10.1016/S0959-437X\(99\)00020-9](https://doi.org/10.1016/S0959-437X(99)00020-9)
- [25] Nick Lane; William F. Martin. The Origin of Membrane Bioenergetics. *Cell* 151(7), 1406–1416 (2012). DOI 10.1016/j.cell.2012.11.050. <https://doi.org/10.1016/j.cell.2012.11.050>
- [26] Jonathan Lombard; Purificación López-García; David Moreira. The Early Evolution of Lipid Membranes and the Three Domains of Life. *Nature Reviews Microbiology* 10, 507–515 (2012). DOI 10.1038/nrmicro2815. <https://doi.org/10.1038/nrmicro2815>
- [27] J A Lake; E Henderson; M Oakes; M W Clark. Eocytes: A New Ribosome Structure Indicates a Kingdom with a Close Relationship to Eukaryotes. *Proceedings of the National Academy of Sciences* 81, 3786–3790 (1984). DOI 10.1073/pnas.81.12.3786. <https://doi.org/10.1073/pnas.81.12.3786>
- [28] Anja Spang; Jimmy H. Saw; Steffen L. Jørgensen; Katarzyna Zaremba-Niedzwiedzka; Joran Martijn; Anders E. Lind; Roel van Eijk; Christa Schleper; Lionel Guy; Thijs J. G. Ettema. Complex Archaea That Bridge the Gap Between Prokaryotes and Eukaryotes. *Nature* 521, 173–179 (2015). DOI 10.1038/nature14447. <https://doi.org/10.1038/nature14447>
- [29] Kalin Vetsigian; Carl Woese; Nigel Goldenfeld. Collective Evolution and the Genetic Code. *Proceedings of the National Academy of Sciences* 103, 10696–10701 (2006). DOI 10.1073/pnas.0603780103. <https://doi.org/10.1073/pnas.0603780103>
- [30] Patrick Forterre. Three RNA Cells for Ribosomal Lineages and Three DNA Viruses to Replicate Their Genomes: A Hypothesis for the Origin of Cellular Domain. *Proceedings of the National Academy of Sciences* 103, 3669–3674 (2006). DOI 10.1073/pnas.0510333103. <https://doi.org/10.1073/pnas.0510333103>

[31] Nicolas Glansdorff; Ying Xu; Bernard Labedan. The Last Universal Common Ancestor: Emergence, Constitution and Genetic Legacy of an Elusive Forerunner. *Biology Direct* 3, 29 (2008). DOI 10.1186/1745-6150-3-29. <https://doi.org/10.1186/1745-6150-3-29>