



## Universal common ancestry, LUCA, and the Tree of Life: three distinct hypotheses about the evolution of life

Joel Velasco<sup>1</sup>

Received: 31 March 2018 / Accepted: 23 August 2018  
© Springer Nature B.V. 2018

### Abstract

Common ancestry is a central feature of the theory of evolution, yet it is not clear what “common ancestry” actually means; nor is it clear how it is related to other terms such as “the Tree of Life” and “the last universal common ancestor”. I argue these terms describe three distinct hypotheses ordered in a logical way: that there is a Tree of Life is a claim about the pattern of evolutionary history, that there is a last universal common ancestor is an ontological claim about the existence of an entity of a specific kind, and that there is universal common ancestry is a claim about a causal pattern in the history of life. With these generalizations in mind, I argue that the existence of a Tree of Life entails a last universal common ancestor, which would entail universal common ancestry, but neither of the converse entailments hold. This allows us to make sense of the debates surrounding the Tree, as well as our lack of knowledge about the last universal common ancestor, while still maintaining the uncontroversial truth of universal common ancestry.

**Keywords** Tree of Life · Last universal common ancestor · LUCA · Common ancestry

### Introduction

What exactly is meant by “universal common ancestry” and why are we so certain that it is true? There is no agreed upon definition, but we can work toward understanding what it means by trying to be explicit about what it entails. As Sober and Steel (2002) note, working biologists usually just assume that common ancestry is true when doing their work. Rarely is it directly questioned. In textbooks, it is

---

✉ Joel Velasco  
joel@joelvelasco.net

<sup>1</sup> Texas Tech University, Lubbock, TX, USA

usually stated that we know that all living things are related because of their shared biochemical features, such as being composed of cells, using DNA, and sharing an identical genetic code.<sup>1</sup> Even if it might seem obvious that these features virtually guarantee the truth of common ancestry via some kind of inference to the best explanation, this inference only makes sense if we have some understanding of what common ancestry actually entails, and citing these shared features is a far cry from actually testing common ancestry directly.

Penny et al. (1982) is sometimes read as an early attempt to try to test directly common ancestry [e.g. by Sober and Steel (2002)]. Penny et al. claim that they are testing “the existence of an evolutionary tree” and claim that it is a prediction of the theory of evolution that “minimal trees with the same taxa should be similar, or ‘congruent’ when constructed from different protein sequences” (197). They construct trees for five different protein sequences in eleven taxa and conclude that the trees are remarkably similar, and thus that there is an evolutionary tree. While their data are just from mammals, the method could in principle be applied to all taxa. Sober and Steel (2002) argue that Penny et al.’s test is flawed on methodological grounds, but whatever evidence this test may or may not have provided, we now know that Penny et al.’s central claim about congruence is empirically false. Different genes quite often have different trees. Much of this incongruence is due to lateral gene transfer. Authors such as Doolittle (1999), who emphasize the importance of lateral gene transfer, sometimes take this to show that there is no Tree of Life (here capitalized following Darwin), yet these same authors are explicit in their acceptance of common ancestry. What they doubt is the *tree structure* of evolutionary history. They do not doubt that there is a shared history (Velasco 2013). Penny et al.’s conclusion that there is an evolutionary tree entails that common ancestry is true, but the denial of the tree does not imply that common ancestry is false. An easy way to see this is that trees have branches that split but never join. But genealogies in general need not be like this and a non-tree-like genealogy can still have a single origin. Sober and Steel (2002) are thus mistaken in treating Penny et al. as attempting to directly test common ancestry.

### What is the tree?

Since the existence of an evolutionary tree is obviously related to common ancestry, the debate about the Tree is a good place to start trying to understand common ancestry. Studies attempting to construct a universal Tree are common, with Hug et al. (2016) providing a convenient, recent example. Hug et al. concatenate data

---

<sup>1</sup> When biologists claim that we know that “all life is related,” this does not require a precise definition of what counts as a living thing. However, it does require several clarifications. First, while viruses may or may not be alive, they are not being considered here. The origins of viruses are not at all clear. While it is possible that they all emerged out of cellular life, it is also possible that some viruses existed before cellular life and may or may not have originated independently of it (Forterre 2016; Nasir et al. 2017). Second, the claim is not necessarily about all life in the universe, but just the life on Earth that we have actually discovered and identified.

from sixteen ribosomal proteins found in 3083 organisms across diverse taxa and find the best fitting tree structure. But building this tree does not really constitute evidence that there is a tree. After all, even if there were no tree, the methods used by Hug et al. would still have produced one. We need a method for testing whether or not there is a tree structure, not a method for producing the best fitting tree.

Penny et al.'s idea for testing the existence of the Tree is a reasonable one. One might think that if the history of species is a branching tree pattern, then the histories of the genes found in organisms of these species should all follow the same tree pattern. If individual species (or perhaps different groups of species) evolved or were created independently, then it would be a highly improbable coincidence if the patterns resulting from forcing the different genes into a tree structure turned out to be the same pattern in every case. But separate ancestry is not the only thing that can cause gene histories to differ. We now know that due to lateral gene transfer, many gene histories differ from the history of species and from each other.<sup>2</sup> Once it is established that different genes have different histories, the argument against the Tree is simple. The Tree is a single pattern that is supposed to represent all of evolutionary history. But different genes have different histories so these histories cannot be represented by a single tree.

Contemporary defenders of the tree do not deny that different genes have different histories. Instead, they argue that this is consistent with the Tree. For example, Galtier and Daubin (2008) argue that since the Tree is a tree of species, this is consistent with any amount of lateral gene transfer. Ciccarelli et al. (2006) argue that the Tree is a tree of organisms, which is what Doolittle and Baptiste (2007) call the "tree of cells". Some defenders even define the tree directly in terms of genetic histories. For example, Puigbò et al. (2009, 2013) argue that the Tree just represents the dominant genetic history or a significant statistical pattern.

Other defenses based on alternative understandings of the Tree are possible. For example, Mindell (2013) argues that the Tree of Life has value as a metaphor, a model, and as a heuristic device, but not as a falsifiable hypothesis. For Mindell, the question is about the value of continuing to use the Tree of Life, not whether there is such a thing. However, if the Tree is a model, then we can just treat the claim "there is a Tree of Life" as the claim that it is a good model. This interpretation seems to capture the debate about the Tree in the literature and to explain the kind of evidence that is brought to bear on the debate.

Whether there is a Tree of Life of course depends on empirical facts about the history of life. But as we can now see, since the Tree is a representation of history, claims about its appropriateness also crucially depend on what aspects of this history we are trying to represent and what kind of representation we are looking for (Velasco 2012). This means that scientific issues as well as the more generally philosophical issues surrounding modeling, idealization, and representation play a major

---

<sup>2</sup> Of course lateral gene transfer is not the only process that causes gene histories to differ. For example, in sexual species, hybridization and incomplete lineage sorting also lead to gene tree incongruence (Maddison 1997).

role. As Doolittle and Bapteste (2007) argue, there are many important patterns in the history of life. Some of these patterns are tree-like and some are not.

I certainly do not intend to settle the debate about the Tree here. But what all parties to the debate agree on is that the claim that there is a Tree of Life is a claim about the *pattern* of evolutionary history and how to appropriately represent that history. Even without being more specific than this, we can already see that the claim about universal common ancestry must be weaker, since the processes of evolution taking place now and the patterns we can use to represent them are relevant to the debate about the Tree, but irrelevant to whether or not common ancestry is true. We can imagine a single origin of life and later, a fully formed functioning cell that gave rise to all of the cells in existence today. This clearly counts as a case where common ancestry is true. But this is consistent with large amounts of lateral gene transfer, hybridization, numerous instances of cell fusions and endosymbiotic events, and many more things that possibly undermine the Tree happening much later. The truth of universal common ancestry (if it is true) does not entail a Tree of Life.

### Back to common ancestry

Since the Tree of Life hypothesis might be false while common ancestry is true, we know that common ancestry has to be logically weaker than the claim that there is a Tree. But this still leaves open exactly what it does entail. A more recent attempt at testing common ancestry is Theobald (2010) who carries out a quantitative test using model selection theory. He specifically asks, “whether the three domains of life (Eukarya, Bacteria and Archaea) are best described by a unified, common genetic relationship (that is, UCA [Universal Common Ancestry] or by multiple groups of genetically unrelated taxa that arose independently and in parallel” (Theobald 2010: 219–220).

Theobald correctly allows that common ancestry is true even if there is no tree structure due to massive lateral gene transfer. However, his definition of common ancestry is too weak. As part of his clarification of the meaning of UCA he says,

“UCA does not demand that the last universal common ancestor was a single organism in accord with the traditional evolutionary view that common ancestors of species are groups, not individuals. Rather, the last universal common ancestor may have comprised a population of organisms with different genotypes that lived in different places at different times” (Theobald 2010: 220).

In order to argue that this definition is too weak, I want to take what first appears to be a slight detour and note that Theobald casually equates the issue of common ancestry with the existence of a common ancestor. Doolittle (2009) and Doolittle and Brunet (2016) have noted the importance of distinguishing between these two ideas, and this distinction introduces yet another complicated wrinkle into our story.

To begin, note that if there are any universal common ancestors, then there is a last (most recent) one. Discussions about the validity of the Tree of Life model as representing a universal phylogeny often discuss the status of the root of the tree (Velasco 2013). The root is usually just thought of as (and called) LUCA—the last

universal common ancestor. There are intense debates about the features of LUCA: for example, about its gene content, how it generated energy, whether it was eubacterial or archaeal, or whether it might have even been protoeukaryotic (Koonin 2003; Ouzounis et al. 2006; Penny and Poole 1999; Weiss et al. 2016; Woese 1998). Theobald is assuming that if common ancestry is true, then there is a LUCA.

But while Theobald is right that fundamentally common ancestry contrasts with separate ancestry, I think it is clear that Theobald has stretched the biological meaning of terms too far. A population whose members live in different places at different times is not a *biological* population at all and is not capable of being the ancestor of anything. Theobald's model effectively treats lateral gene transfer the same as parent–offspring relationships for the purposes of identifying whether two groups are related. He then makes matters worse by allowing that the “ancestor” is simply a set of organisms that do not have to be united in any special ontological way. But suppose that two groups originated and evolved separately and then some of their descendants began to share genes by lateral transfer. Surely, in this scenario, there is no LUCA. Questions such as whether LUCA was autotrophic make no sense if LUCA could consist of an arbitrary set of organisms that have very different properties.

### What is LUCA?

Just as we need to clarify what is meant by “common ancestry” we also need to clarify what is meant by “LUCA” and in particular, what a common ancestor is. At a minimum it has to be capable of being an ancestor and that rules out very loose understandings such as Theobald's. But many different kinds of entities are ancestors. For example, Sober and Steel (2002: 401–402) describe common ancestry in the following way:

The Hypothesis of Common Ancestry says that there exists a single ancestral origin to which all present-day living things trace back. This hypothesis competes with alternative hypotheses that say that the number of ancestors is 2, 3, ..., or  $n$ . More precisely, the hypotheses we want to consider have the following form for  $i = 1, 2, \dots, n$ :

CA- $i$  There existed a set  $A$  consisting of  $i$  species, and no set with fewer than  $i$  species, such that:

- (i) none of the species in  $A$  are ancestral to any other species in  $A$ ,
- (ii) each of the current species ( $S_1, S_2, \dots, S_n$ ) has at least one ancestor in  $A$ , and
- (iii) each species in  $A$  is ancestral to at least one  $S_k$ .

Using their notation, we could say that there is a LUCA iff CA-1 is true. On this reading, LUCA would be a single species. Here the “universal” refers to the fact that this species is an ancestor of all present day species. But “species” may not be the right category of entity. It is not entirely clear what a species is or if there even are

species (especially in prokaryotes), and the category is especially problematic for very early life (Ereshefsky 2010; Doolittle 2012).

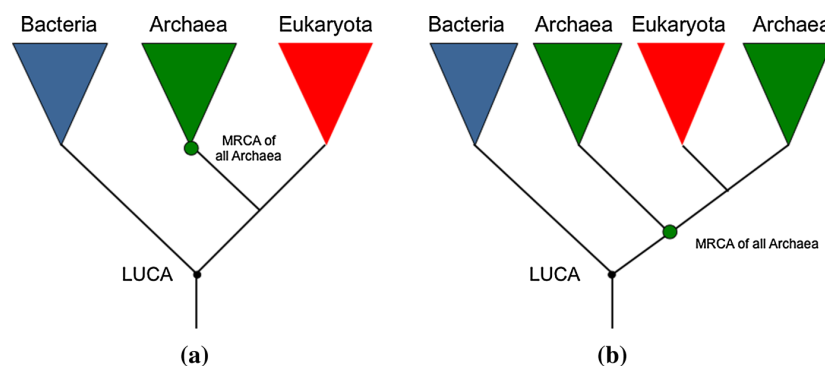
Sober (2008: 271) keeps the central idea of a single ancestral origin, but changes these clauses by replacing “species” with “organism”. This eliminates the worry that a species might not be an appropriate entity as an ancestor. But the organism reading has similar problems in that it is also unclear what an organism is and again, this is especially true for very early life.<sup>3</sup> As a first pass, other reasonable candidates for a LUCA are a population, a cell, or a genome, any of which can be ancestors.

To see how these different options for the ontology of LUCA might play out, we can examine a few empirically reasonable models of what the early history of life might have looked like. A natural question to ask, as Theobald does, is how the three domains of life are related. Assuming a standard model of cellular reproduction by binary fission, one possibility is that each of these domains is a monophyletic group of cells. In that case, the relevant question is how the three domains are related. However, it may be that one of these domains emerged from within another one making that latter domain paraphyletic. It could even be that none of the domains are monophyletic.

If we attempted to build a universal phylogenetic tree, then determining which of these hypotheses was correct requires determining the location of the root of the tree. But a major difficulty in locating the root of a universal Tree as opposed to determining the root of any other phylogenetic tree is that to root trees, we need to determine character polarity, which means inferring the ancestral state of the trait. The standard way to do this is by using an outgroup closely related to the groups you are considering. But of course there is no outgroup for all of life. Clever methods, such as using one version of an ancient gene duplication as the outgroup for the other universal gene, were among the first studies to produce well-supported results. Gogarten et al. (1989) and Iwabe et al. (1989) independently used this method with different genes to arrive at the conclusion that life divides into two great clades—the Bacteria on the one hand, and a second branch consisting of the Archaea and the Eukaryotes. The iconic universal tree of Woese et al. (1990) created using genetic distances in the 16S/18S SSU rRNA molecule used these studies to determine its root. Later studies, such as Baldauf et al. (1996), respond to criticisms of the original studies and greatly expand the data set to reach the same conclusions. A schematic three-domain tree with this suggested placement of the root is depicted in Fig. 1a.

However, a number of more recent studies have strongly indicated that the Eukaryotes arose from within the Archaea—probably as the result of some kind of fusion of an archaeon with a bacterium (Williams et al. 2013; Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017). This leads to a two-domain model of the evolution of life, which solves the problem of the order of the branching between domains, but still leaves open just what the ancestor was like. A schematic two-domain tree is depicted in Fig. 1b.

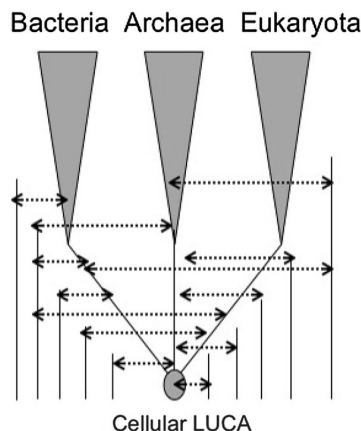
<sup>3</sup> Clarke (2010) lists thirteen different candidate definitions of individual or organism that have been defended in the literature.



**Fig. 1** Schematic diagrams of a three-domain **(a)** and two-domain **(b)** tree. “LUCA” refers to the last universal common ancestor while the “MRCA” of all Archaea is the most recent common ancestor of the Archaea. In **a** each of the three domains is monophyletic. In **b**, the MRCA of the Archaea is an ancestor of all of the Eukaryotes as well making the Archaean domain paraphyletic. Thus some Archaea are more closely related to the Eukaryotes than they are to some other Archaea

On either the two-domain or the three-domain account we have the same question—what does the last common ancestor of Bacteria and Archaea/Eukarya look like? A naive view is that LUCA is just a single-celled organism that most likely had the traits common to all of the domains of life. But this basic thought leads to an absurdity: LUCA contained genes for nearly all types of biochemical reactions known in Bacteria and Archaea today and therefore had a genome larger than any known prokaryotic genome today. This is what Doolittle et al. (2003: 46) termed “the genome of Eden”. Lateral gene transfer can explain why genes are often found in many distantly related taxa, but it does not solve the fundamental problem of what the source of all these genes are if all of these genes trace back to a single ancestral genome. Doolittle (2005, 2009) argues against the existence of LUCA (on this way of understanding it), saying, “We (some of us) do doubt that there ever was a single universal common ancestor (a last universal common ancestor or LUCA), if by that is meant a single cell whose genome harboured predecessors of all the genes to be found in all the genomes of all cells alive today” (Doolittle 2009: 2221).

No biologist actually believes that there ever was a single cell that contained ancestral copies of every gene in existence today (though perhaps some models that biologists use accept this implicitly). No defender of LUCA has this in mind. For a more empirically plausible story, consider a scenario described by Zhaxybayeva and Gogarten (2004) which Becerra et al. (2007) later argue is likely correct: namely, that there was a single cellular organism LUCA that is the cellular ancestor of all cells today, but that this cell was not alone at the time. Other cells in other lineages existed as well and contributed genes to LUCA’s lineage via lateral transfer. These other lineages have since gone extinct leaving only descendants of LUCA still alive. In this picture, CA-1 is true in the sense that there is a single cell (organism) that is an ancestor of all living cells (organisms). But there is no single ancestral genome that was the ancestor of all living genes. Allowing



**Fig. 2** One reasonable model of the phylogeny of life with a LUCA. Here LUCA is a single cell that is the ancestor of all cells today. The horizontal dotted lines represent lateral gene transfers between lineages. In this model, some genes in living cells do not descend from any genes in LUCA—rather, they descend from genes that were laterally transferred from branches of life that later went extinct. The three-way split after LUCA is not an assertion that there are actually three separate lineages here, but is meant to be consistent with any possible ordering of the branches. This figure is inspired by Andam and Gogarten (2011)

for lateral gene transfer in our models gives us one way of understanding LUCA that is much more plausible than the genome of Eden model. This model is depicted in Fig. 2.

In the model depicted in Fig. 2, there is a single cell that is the ancestor of all cells today and so could arguably be called LUCA. But this is not the only empirically reasonable story about the history of life. Over the course of much of his career, Carl Woese and a number of coauthors in various publications put forward a story that is, in many people’s minds, inconsistent with the existence of LUCA.

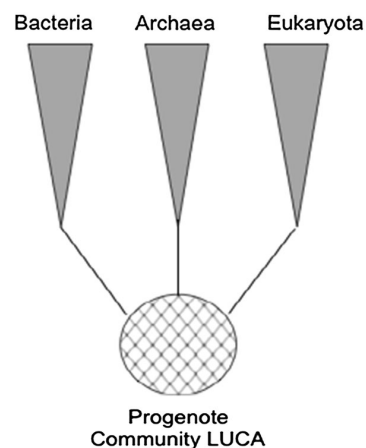
Woese (1998) argues that,

“[T]he ancestor cannot have been a particular organism, a single organismal lineage. It was communal, a loosely knit, diverse conglomeration of primitive cells that evolved as a unit, and it eventually developed to a stage where it broke into several distinct communities, which in their turn became the three primary lines of descent. The primary lines, however, were not conventional lineages. Each represented a progressive consolidation of the corresponding community into a smaller number of more complex cell types, which ultimately developed into the ancestor(s) of that organismal domain. The universal ancestor is not an entity, not a thing. It is a process characteristic of a particular evolutionary stage” (Woese 1998: 6858).

Sober (2008) treats this as claiming that there was no LUCA because he equates LUCA with an organism. Organisms can count as ancestors. Loosely knit communities of protocells cannot. So if Woese’s story is correct, there is no LUCA. Doolittle agrees. Regarding the community as ancestor theory, he says, “Indeed, I too think



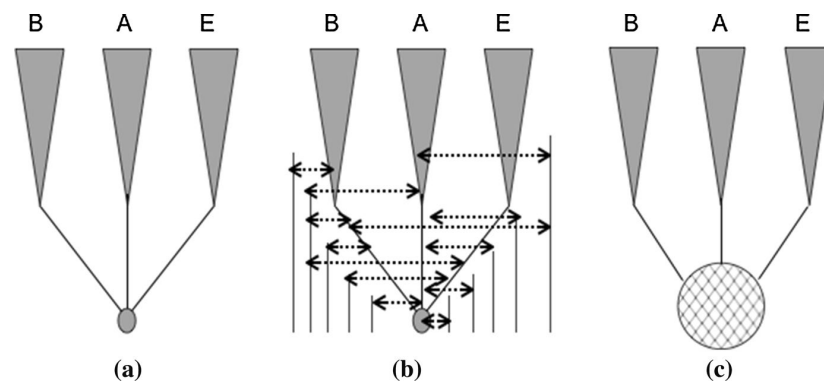
**Fig. 3** A schematic depiction of the Woese model with three separate lines of descent emerging from the progenote phase of life. Here “LUCA” is the entire community of progenotes



that ‘LUCA was a population,’ but I argue that this in fact means that there actually was no LUCA” (Doolittle 2005: 125).

The Woese model arises from thinking about the evolution of the cell itself and in particular, the evolution of translation. Long before cells had the advanced translation mechanisms of today, cells or pre-cellular entities of some type must have had a rudimentary translation apparatus that was less complex and more prone to error and unable to produce many of the modern, complex proteins. Woese and Fox (1977) coined the word “progenote” for such entities. This term is meant to mark the idea that the genotype–phenotype link had yet to complete its evolution. Gorgarten and Olendzenski (1999) point out that numerous authors have mistakenly identified “progenote” as simply another name for LUCA, but this usage is incorrect. The term “progenote” was introduced as a name for a specific type of entity. The idea of the progenote and how it relates to common ancestry was then further developed in Woese (1983, 1987, 1998).

Not only are progenotes simple entities utilizing only relatively few small proteins, but because they lacked cell walls, their parts would freely move in and out between other progenotes in their immediate environment in a process which has similar results to lateral gene transfer. This means that individual progenotes are not really individuals in a biological sense at all. Rather, clusters of progenotes evolve together like overlapping communities without forming genealogical lineages. Woese argues that the process is analogous to a kind of physical annealing where the progenote world undergoes rapid evolutionary change until eventually we reach the emergence of cellular structures that are more stable. Once translation becomes accurate enough, communities make the transition across what Woese et al. call the “Darwinian threshold.” After this, the idea of genealogy makes sense and individual organism-like entities begin to form true lineages. While lateral gene transfer can still occur, vertical evolution is now in play, and the concept of a phylogenetic tree makes sense. What the progenote model of LUCA asserts is that this threshold was crossed independently at different times. Woese (1998) postulates that each of the three primary lines



**Fig. 4** Three possible models for LUCA. For simplicity, all three models depict the three domains (labeled B, A, and E) as monophyletic and leave open the branching order between them. In **a** there is a single genomic ancestor of all extant life. This would clearly count as LUCA, however, it is not empirically reasonable. On the other hand, **b** and **c** are empirically reasonable models. In **b** there is a single cellular ancestor of all cells, but it is not a genomic LUCA. In **c** there is a community style ancestor. However, the properties of this community may make it ontologically unsuitable as a single biological entity and so therefore unsuitable as LUCA. **a** Genome of Eden. **b** Cellular LUCA. **c** Progenote community LUCA

of descent (Bacteria, Archaea, and Eukaryotes) crossed the threshold independently. A schematic representation of the Woese model is depicted in Fig. 3.

In Figs. 2 and 3, we see these two distinct types of models for LUCA. We can think of an organismal LUCA model as asserting that LUCA was a cell whose ancestors had already crossed the Darwinian threshold, while the Woese community-style LUCA model asserts that LUCA was a community of pre-cellular (and pre-organismal) progenotes. This cell/pre-cell distinction marks an important difference in views about LUCA. For example, Patrick Forterre and his coauthors in numerous papers refer to “LUCA (the Last Universal Cellular Ancestor)” [e.g. Philippe and Forterre (1999), Forterre et al. (2004), Forterre (2010)]. One reason for this name is that we might want to consider viruses as living, and if we do, then we want to be clear that LUCA might just be the ancestor of all cellular life and not necessarily the ancestor of *all* life. But it is also clear that these authors assume that the ancestor of cellular life must itself be cellular. Obviously a pre-cellular community cannot count as a LUCA in this sense! Syvanen (2002: 265) explicitly denies the existence of a LUCA because “the modern cell could have evolved in multiple parallel lineages.” But note that in both Figs. 2 and 3, cells evolve in multiple lineages. Both are empirically reasonable models and show that there is a LUCA of a certain type and not of another type. These two models of LUCA, along with our original Genome of Eden model are summarized in Fig. 4 below.

As with our debate about the Tree of Life, there is both an empirical and a semantic component to the debate about LUCA. Whether or not there was a LUCA depends both on the empirical facts, many of which are unknown at this time, and also what you mean by “LUCA” in the first place. The basic, neutral idea is that there is a single universal ancestor. This is CA-1 in the sense of Sober and Steel, but

just without specifying what the “1” has to be. Some authors think that LUCA has to be an organism—a single cell. Others are more permissive, allowing that LUCA could be a population. This is reasonable as long as it is an integrated, localized, biological population. Yet despite this ambiguity, as in the case of common ancestry, it is clear that the existence of a LUCA does not imply that there is a Tree of Life. To repeat, the Tree essentially involves representing evolutionary processes that happen much later than the time of LUCA. On the other hand, the converse implication is correct. On any reasonable understanding of the Tree of Life, that Tree will have a root and that root is appropriately called LUCA.<sup>4</sup>

### Back (again) to common ancestry

So we have established that the Tree of Life entails a LUCA, but not vice versa. But what about the relationship between LUCA and common ancestry? If there is a LUCA, then universal common ancestry is most certainly true. But for some ways of understanding LUCA, the converse is not true. We can see that this must be the case by looking at particular views such as Doolittle's. Doolittle and Brunet (2016) criticize community-style LUCA views such as Woese (1998) and Arnoldt et al. (2015) for conflating “having common ancestry” with “having a common ancestor”. In responding to Jonathan Wells' (2006) use of Doolittle's work (among others) to attack common ancestry, Doolittle responds that Wells is conflating the existence of a certain kind of common ancestor with universal common ancestry. He then follows this up by saying, “That ‘common ancestry’ does not entail a ‘common ancestor’ is perhaps a subtle point—one that Woese and co-authors recently explicated very nicely in the case of the genetic code by the way (Vestigian [sic] et al. 2006)” Doolittle (2009: 2221).

Vetsigian et al. (2006) defend a “communal evolution” model of the origin of the genetic code. What Doolittle is arguing is that this kind of loose-knit community of pre-cellular entities sharing traits through a kind of lateral gene transfer is not a single entity with the right kind of ontological status to call an “ancestor” of anything. For example, though Doolittle does not put it this way, it could easily be argued that if the concept of reproduction does not apply, then we do not have genealogical relationships and so *a fortiori* no ancestry relationships. So while we might call this community a “population”, the entities inside it cannot be ancestors of anything, nor can the collective as a whole. Yet at the same time, Doolittle is claiming that in this scenario it makes sense to talk about common ancestry. Why?

Common ancestry is so important in biology because of its role in explaining the distributions of traits across the diversity of life. To take the classic case, why do the

---

<sup>4</sup> In response to my claim that the Tree of Life must have a root, a number of audience members have pointed out at various talks that there are rooted and unrooted phylogenetic trees. But the Tree of Life is supposed to be a representation of evolutionary history and as such it must have a root. To see this, it is perhaps enough to note that on an unrooted phylogenetic tree, there is no distinction between a “clade” and its complement (all the tips not in that clade). To ensure that collections such as the “non-mammals” (which includes everything from lizards, the mosses, and *E. coli*) are not real groups requires a root.

limbs of frogs, lizards, birds, and mammals (including whales, bats, and humans) all have the same bone structures consisting of a humerus, radius, ulna, carpals, metacarpals, and phalanges? Because the limbs are *homologues*—they have all evolved from a common ancestral form of the tetrapod limb. The fact that the tetrapods all descended from a common ancestor is of course essential to this common cause explanation. Sober and Steel (2002: 401) say, “The Hypothesis of Common Ancestry says that there exists a single ancestral origin to which all present-day living things trace back.” This is exactly right and no further elaboration of “origin” is needed or desirable. They go on to point out that by “single ancestral origin” they don’t mean to imply that life originated only once. Only that all but one branch must have died out or there was some sort of bottleneck that all life can trace back to. “Origin” here is vague, but importantly so. If there were a single organism, population, or species ancestral to all life, common ancestry would clearly be true. But there are other possibilities. The major branches of life could have emerged from a common physical collective such as in the Woese et al. story. Recall that Woese (1998: 6858) claims that the ancestor “evolved as a unit.” This is the key to understanding why it counts as common ancestry—it allows that universal traits such as the genetic code evolved only once.<sup>5</sup>

Admittedly, some understandings of LUCA are so loose that they are effectively just equivalent to common ancestry. For example, Weiss et al. (2016: 1) begin their paper by saying, “The last universal common ancestor (LUCA) is an inferred evolutionary intermediate that links the abiotic phase of Earth’s history with the first traces of microbial life in rocks that are 3.8–3.5 billion years of age.” Here, whatever the intermediate is, it qualifies as LUCA. Weiss et al. (2016) suggest that LUCA inhabited hydrothermal vents and had a primitive metabolism lacking the genes for many types of essential biosynthesis, instead relying on geochemical reactions in its environment. It is thus only “half-alive.” This may be another case of a LUCA that doesn’t deserve the name. But even if this is true, this is a clear example of common ancestry. Hundreds of genes were present as was a genetic code. We can attribute specific properties to this “LUCA” and changes in its properties would affect its “descendants” (later stages of the processes going on in the vents). Assuming the Weiss et al. scenario were true, then if we asked about the origin of the genetic code, we could tell a story about it having a single origin in the vents.<sup>6</sup>

<sup>5</sup> A different story might be told utilizing the model of Doolittle and Booth’s (2017) “it’s the song not the singer” idea. Here, we could imagine a community where a song (a pattern of interactions such as metabolic, structural, or developmental interactions) gets preserved, replicated, and altered over time without the underlying physical lineages necessarily being the ancestors of the future lineages singing the song. Different singers (different taxa) can be recruited horizontally at different points during evolutionary history. This might also be called common ancestry without a LUCA.

<sup>6</sup> A reader here may worry about the definition of “life” and how that is being used in the argument. If we want to answer the question “How many times did life begin?” we obviously need to know what qualifies as life. That problem is extremely tricky and may not even have a correct answer at all. But we do not have to answer the “what is life” question in order to answer the question of whether a given collection of entities has a common origin. For example, whether the LUCA in Weiss et al.’s (2016) story counts as alive or not is irrelevant since it is a causal bottleneck in the sense that living things can causally trace their properties back to the properties of LUCA.

Perhaps the easiest way to understand the implications of universal common ancestry is to see what kinds of scenarios would render it false. Weiss et al.'s investigation of the properties of LUCA presupposes that there is a single intermediate. If one abiotic intermediate led to the Bacteria and another, different abiotic intermediate led to the Archaea, we cannot simply take the two together and collectively call them "LUCA". That would be to commit Theobald's mistake. Individuating intermediate stages is *in general* an extremely tricky problem. But in this case, the purpose of the individuation is to distinguish *common* ancestry from *separate* or *independent* ancestry. This can be done in an explanatory framework. If the explanation for why the genetic code is the same in Archaea and Bacteria involves the fact that they evolved from the same entity, then that is common ancestry. If the code evolved once in the lineage leading to the Archaea, evolved independently in the lineage leading to the Bacteria, and the similarity is just a coincidence, then obviously this is separate ancestry. But importantly, separate ancestry does not entail that the similarities have to be coincidental. If there are stereochemical reasons why various features of the code have to be the way they are, this would of course explain why the code is similar in this way in different organisms. But if the code evolved twice in this same way, that is separate, not common, ancestry. If a kind of natural selection led the code to evolve into the same optimal state in both lineages, again, that is separate ancestry. These may count as cases where the code in Archaea and Bacteria has the same *type* of origin, but they aren't cases where they have the same *token* origin. Hence the code would have more than one origin and thus separate ancestry.

Of course the genetic code may just happen to be an instance of a trait that evolved twice even though the common ancestry of life is true. Numerous traits have evolved more than once in the history of life: morphological features such as eyes and wings have evolved many times over in animals (Land and Nilsson 2002; Alexander 2015) while chemical pathways such as those leading to C<sub>4</sub> photosynthesis independently evolved over 45 times in 19 families of angiosperms (Sage 2004). Likewise, we know that proteases, which are enzymes that break down proteins and can be found in all living organisms, have multiple origins as well (Buller and Townsend 2013). Given all these examples, the interesting question is whether there are any universal traits that have evolved only once.

The reason that universal common ancestry is so important to our understanding of biology is not that all traits have evolved only one time in evolutionary history. It is that common ancestry *allows* a single origin explanation for any number of universal traits. Shared, contingent features of all living things—such as genetic information encoded as DNA, similarities in biochemical processes such as transcription and translation, and the genetic code for translating genetic information into proteins—all point to a single, shared origin. It may turn out that some of these features could have evolved more than once. But separate ancestry would necessitate that they all did and that is what seems so incredible.

To be fair, and to expose even further ambiguities in the literature, I should point out that there is a way that a trait could be universally distributed, have a single evolutionary origin, and yet for life itself to be the product of separate ancestry. Lateral gene transfer makes this kind of explanation possible in principle. Theobald's use of LUCA as an arbitrary set of organisms whose descendants share genes is clearly inappropriate.

But his idea that lineages sharing genes via lateral transfer “become one species” and so can count as an example of common ancestry is not obviously wrong. This idea also explains Syvanen’s (2005) claim that we can have common descent without a common ancestor. It is tempting to say that if all of life shares numerous, universally distributed genes, then because those genes each have a single origin and have been transferred throughout the rest of life via lateral transfer, this process should count as a kind of common descent. Syvanen (2005) and Theobald (2010) both explicitly argue that it should.

Here, I disagree. The common descent of different branches of life *explains* why distinct lineages can share genes in the first place. The basic biochemistry of living things has to already be similar enough in fundamental ways in order for lateral gene transfer to be possible. But the lateral transfer of genes does not itself *constitute* the common descent of the donor and recipient organisms. I would argue that if life started more than once in a separate place and time, evolved independently, and then, without ever forming any kind of reproductive or community bottleneck, the distinct lineages began to share genes via lateral transfer, this should count as separate ancestry.

Imagine how our fundamental understanding of biology would have to change if we were to discover that this scenario were true. We would be searching for the explanation as to why distinct branches of life, which evolved separately, turned out to be so fundamentally compatible. This discovery would be the equivalent to discovering life that evolved on another planet and that was biochemically compatible with life on our planet. It is certainly conceptually possible. Nothing we know about the nature of living things absolutely rules it out. But even without any direct tests of the truth of the common ancestry of all life on earth today, what we think we know about the contingency of the outcomes of the evolutionary process ensures that this would be a shocking development. If life evolved independently on another planet and turned out to be made of cells, to use DNA to store genetic information, and have a genetic code similar to ours, it would be absolutely astonishing. Nothing about our current understanding of the physics and chemistry of life could explain why the biology of these other life forms was so similar to ours. So why are we *not* shocked to find out that a newly discovered species on earth does have all of these traits? Because we *do* know the explanation for these similarities: common descent. It is this explanatory feature of universal common ancestry that is crucial to understanding the biology of life on this planet.

## Conclusion

So to briefly summarize, is there a universal Tree of Life? The answer is not clear. It depends both on what you mean by “Tree of Life” and on the empirical facts about the evolution of early life and about life’s evolution over the last few billion years. This later evolution includes facts about the frequency of various types of events like lateral gene transfer, hybridization, and endosymbiosis. It also matters what the purpose of the Tree is and what exactly it is supposed to represent. If the Tree is supposed to be a single tree representing every aspect of the evolution of all life forms, then there is no such thing. But if the model only has to be relevantly tree-like and

only has to be represent some important features of evolutionary history, then it is much more reasonable to say that there is a Tree.

Is there a LUCA—a last universal common ancestor? Here again, empirical facts about what the early evolution of life was like are very important and many of the details are currently unknown. But one thing that is clear is that there never was a single cellular organism of the kind that exists today that alone gave rise to all of the current diversity of life. There may have been a single cellular ancestor of all cells today that might appropriately be called a LUCA, but many genes and gene families cannot trace their ancestry to that cell's genome. Another possibility is that there may have been a population of pre-cellular entities that collectively led to multiple cellular lineages. But whether this population is appropriately called LUCA depends on its ontological status as well as a semantic question about the appropriateness of the application of terms like “ancestor”.

What about universal common ancestry? This is almost certainly true. Shared, contingent features of all living things such as their basic biochemistry as well as the singular origin of the genetic code all point to a single, shared origin of all life.

In this paper, I have not attempted to argue for any particular view about the Tree of Life or about LUCA. I have, however, argued that the Tree is a claim about the pattern of evolutionary history and that LUCA is a claim that some specific type of biological entity that existed in the past and is the ancestor of all extant life. But what that pattern has to be like and what that entity has to be like, I leave open for future debate.

Universal common ancestry is a bit different. I have argued for a particular understanding of universal common ancestry, namely, that all living things trace their causal history back to a single, token origin. While this common cause structure seems quite general and vague, we should *not* make it more precise. The truth of common ancestry leaves open the physical underpinnings of the nature of this common cause. Common Ancestry as an explanatory strategy can only be successful if a common cause actually existed. But what the nature of that cause was is not important.

I hope to have made clear why it is important to keep these three hypotheses separate, though we should recognize the logical relationships between them. The existence of the Tree would entail the existence of LUCA, which would entail universal common ancestry. However, neither of the converse entailments holds. It is reasonable to accept that there is a Tree of Life, that this tree has a root that we can call LUCA, and that all extant life on this planet is related through common ancestry. But it is perfectly possible to accept universal common ancestry in the sense of a single token common cause origin of extant life today while denying that there is a single entity of a specific ontological kind—a universal common ancestor. And even if one does accept that there is a universal common ancestor (and so a most recent one, LUCA), it is perfectly reasonable to deny that there is a universal Tree of Life in the sense of a single branching tree diagram representing the evolutionary history of life.

**Acknowledgements** I would like to thank Ford Doolittle, Maureen O'Malley, Jeremy Schwartz, Elliott Sober, Quayshawn Spencer, several anonymous referees, and audiences at the University of Pennsylvania and the 2016 Philosophy of Science Association meeting for helpful comments and discussion.

## References

- Alexander DE (2015) *On the wing: insects, pterosaurs, birds, bats and the evolution of animal flight*. Oxford University Press, Oxford
- Andam CP, Gogarten JP (2011) Biased gene transfer in microbial evolution. *Nat Rev Microbiol* 9:543–555
- Arnoldt H, Strogatz SH, Timme M (2015) Toward the darwinian transition: switching between distributed and speciated states in a simple model of early life. *Phys Rev E* 92:052909
- Baldauf SL, Palmer JD, Doolittle WF (1996) The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc Natl Acad Sci USA* 93(15):7749–7754
- Becerra A, Delaye L, Islas S, Lazcano A (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(1):361–379
- Buller AR, Townsend CA (2013) Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proc Natl Acad Sci USA* 110(8):E653–E661. <https://doi.org/10.1073/pnas.1221050110>
- Ciccarelli FD, Doerks T, von Mering C, Creevey C, Snel B, Bork P (2006) Towards automatic reconstruction of a highly resolved Tree of Life. *Science* 311:1283–1287. <https://doi.org/10.1126/science.1123061>
- Clarke E (2010) The problem of biological individuality. *Biol Theor* 5(4):312–325
- Doolittle WF (1999) Phylogenetic classification and the universal tree. *Science* 284(5423):2124–2128
- Doolittle WF (2005) If the Tree of Life fell, would we recognize the sound? In: Sapp J (ed) *Microbial phylogeny and evolution: concepts and controversies*. Oxford University Press, Oxford, pp 119–133
- Doolittle WF (2009) The practice of classification and the theory of evolution, and what the demise of Charles Darwin's Tree of Life hypothesis means for both of them. *Philos Trans R Soc B Biol Sci* 364(1527):2221–2228
- Doolittle WF (2012) Population genomics: how bacterial species form and why they don't exist. *Curr Biol* 22(11):R451–R453. <https://doi.org/10.1016/j.cub.2012.04.034>
- Doolittle WF, Bapteste E (2007) Pattern pluralism and the Tree of Life hypothesis. *Proc Acad Natl Sci USA* 104:2043–2049
- Doolittle WF, Booth A (2017) It's the song, not the singer: an exploration of holobiosis and evolutionary theory. *Biol Philos* 32:5–24
- Doolittle WF, Brunet TDP (2016) What is the Tree of Life? *PLoS Genet* 12(4):e1005912. <https://doi.org/10.1371/journal.pgen.1005912>
- Doolittle WF, Boucher Y, Nesbø CL, Douady CJ, Andersson JO, Roger AJ (2003) How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos Trans B* 358(1429):39–58
- Ereshefsky M (2010) Microbiology and the species problem. *Biol Philos* 25:67–79
- Forterre P, Filée, J, Myllykallio H (2004) Origin and evolution of DNA and DNA replication machineries. In: Ribas de Pouplana L (ed) *The genetic code and the origin of life*. Landes Bioscience, Austin, pp 145–168
- Forterre P (2010) Defining life: the virus viewpoint. *Orig Life Evol Biosph* 40(2):151–160. <https://doi.org/10.1007/s11084-010-9194-1>
- Forterre P (2016) To be or not to be alive: how recent discoveries challenge the traditional definitions of viruses and life. *Stud Hist Philos Biol Biomed Sci* 59:100–108
- Galtier N, Daubin V (2008) Dealing with incongruence in phylogenetic analyses. *Philos Trans R Soc Lond B Biol Sci* 363:4023–4029
- Gogarten JP, Olendzenski L (1999) The progenote. In: Creighton T (ed) *Encyclopedia of molecular biology*. Wiley, NY
- Gogarten JP, Kibak H, Dittrich P, Taiz L, Bowman EJ, Bowman BJ, Manolson MF et al (1989) Evolution of the vacuolar H<sup>+</sup>-ATPase: implications for the origin of eukaryotes. *Proc Natl Acad Sci USA* 86(17):6661–6665



- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hermsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF (2016) A new view of the Tree of Life. *Nat Microbiol*. <https://doi.org/10.1038/nmicrobiol.2016.48>
- Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci USA* 86(23):9355–9359
- Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* 1(2):127–136
- Land MF, Nilsson D-E (2002) *Animal eyes*. Oxford University Press, Oxford
- Maddison W (1997) Gene trees in species trees. *Syst Biol* 46(3):523–536
- Mindell DP (2013) The Tree of Life: metaphor, model, and heuristic device. *Syst Biol* 62(3):479–489
- Nasir A, Kim KM, Caetano-Anollés G (2017) Phylogenetic tracings of proteome size support the gradual accretion of protein structural domains and the early origin of viruses from primordial cells. *Front Microbiol* 8:1178. <https://doi.org/10.3389/fmicb.2017.01178>
- Ouzounis CA, Kunin V, Darzentas N, Goldovsky L (2006) A minimal estimate for the gene content of the last universal common ancestor—exobiology from a terrestrial perspective. *Res Microbiol* 157:57–68
- Penny D, Poole A (1999) The nature of the last universal common ancestor. *Curr Opin Genet Dev* 9(6):672–677
- Penny D, Foulds LR, Hendy MD (1982) Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature* 297(5863):197–200
- Philippe H, Forterre P (1999) The rooting of the universal Tree of Life is not reliable. *J Mol Evol* 49:509–523
- Puigbò P, Wolf YI, Koonin EV (2009) Search for a Tree of Life in the thicket of the phylogenetic forest. *J Biol* 8:59. <https://doi.org/10.1186/jbiol159> PMID: 19594957
- Puigbò P, Wolf YI, Koonin EV (2013) Seeing the Tree of Life behind the phylogenetic forest. *BMC Biol* 11:46. <https://doi.org/10.1186/1741-7007-11-46> PMID: 23587361
- Sage RF (2004) The evolution of C<sub>4</sub> photosynthesis. *New Phytol* 161:341–370. <https://doi.org/10.1111/j.1469-8137.2004.00974.x>
- Sober E (2008) *Evidence and evolution: the logic behind the science*. Cambridge University Press, Cambridge
- Sober E, Steel M (2002) Testing the hypothesis of common ancestry. *J Theor Biol* 218(4):395–408
- Spang A, Saw JH, Jorgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJ (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521:173–179
- Syvanen M (2002) On the occurrence of horizontal gene transfer among an arbitrarily chosen group of 26 genes. *J Mol Evol* 54:258–266. <https://doi.org/10.1007/s0023901-0007-z>
- Syvanen M (2005) The last universal common ancestor. The panda's thumb. <https://pandasthumb.org/archives/2005/08/the-last-univer.html>. Accessed 30 August 2005
- Theobald DL (2010) A formal test of the theory of universal common ancestry. *Nature* 465:219–222
- Velasco JD (2012) The future of systematics: tree thinking without the tree. *Philos Sci* 79:624–636
- Velasco JD (2013) The Tree of Life. In: Ruse M (ed) *The Cambridge encyclopedia of darwin and evolutionary thought*. Cambridge University Press, Cambridge, pp 340–345
- Vetsigian K, Woese C, Goldenfeld N (2006) Collective evolution and the genetic code. *Proc Natl Acad Sci* 103(28):10696–10701
- Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, Martin WF (2016) The physiology and habitat of the last universal common ancestor. *Nat Microbiol*, vol 1. <https://doi.org/10.1038/nmicrobiol.2016.116>
- Wells J (2006) *The politically incorrect guide to darwinism and intelligent design*. Regnery Publishing Inc, Washington, DC
- Williams TA, Foster PG, Cox CJ, Embley TM (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504:231–236
- Woese CR (1983) The primary lines of descent and the universal ancestor. In: Bendall DS (ed) *Evolution from molecules to men*. Cambridge University Press, Cambridge, pp 209–233
- Woese CR (1987) Bacterial evolution. *Microbiol Mol Biol Rev* 51(2):221–271
- Woese C (1998) The universal ancestor. *Proc Acad Natl Sci USA* 95(12):6854–6859

- Woese CR, Fox GE (1977) The concept of cellular evolution. *J Mol Evol* 10(1):1–6. <https://doi.org/10.1007/BF01796132>
- Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Acad Natl Sci USA* 87(12):4576–4579
- Zaremba-Niedzwiedzka K, Careres EF, Saw JH, Backstrom D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, Stott MB, Nunoura T, Banfield JF, Schramm A, Baker BJ, Spang A, Ettema TJ (2017) Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541:353–358
- Zhaxybayeva O, Gogarten P (2004) Cladogenesis, coalescence and the evolution of the three domains of life. *Trends Genet* 20(4):182–187