Evaluation of the Genetic Code as Evidence For Evolution

Cornelius Hunter, Ph.D.
William Jessup University
chunter@jessup.edu

Abstract

Evolution is considered to be not merely a theory but also a fact, due to a range of evidence. One of the most celebrated evidences for evolution, over the past half century, is the genetic code. Today this evidential argument is widely accepted and often asserted without justification. However, scientific findings in recent decades, as well as erroneous claims from the beginning, severely undercut this evidential argument. Here I summarize the affirmative case for the argument that the genetic code is powerful evidence for evolution, as well as its failures. I conclude that we have erred, and as research continues into the origin of the genetic code, we also need to present the empirical evidence accurately and in an unbiased manner.

Keywords: Genetic code, evolution, common descent, Crick

Introduction

Evolution is considered to be both a theory and a fact. It is considered to be a theory in the sense that researchers continue to investigate how evolution is thought to have occurred. It is considered to be a fact in the sense that, aside from uncertainties of how it occurred, certain evidence is thought to reveal that evolution, broadly construed, did occur, beyond any reasonable doubt. (Baum et al. 2016) One of those evidences, ever since its discovery in the mid twentieth century, is the genetic code.

The genetic code is used inside living cells to synthesize proteins. The protein synthesis process includes four major steps: transcription, editing, translation, and recharging. Transcription makes a ribonucleic acid (RNA) copy of a region of the deoxyribonucleic acid (DNA) molecule. The resulting copy, which consists of a string of hundreds or thousands of nucleotides, may be edited. Then, in the translation process, the edited copy is read, three nucleotides at a time. These nucleotide triplets are referred to as *codons*. At one end of a transfer RNA (tRNA) molecule three nucleotides protrude outward, forming an *anticodon*. Anticodons and codons bind, and at the opposing end of the tRNA molecule is an amino acid. There are, nominally, 20 different types of amino acids used in synthesizing proteins. In the translation process, tRNA molecules bind, in turn, to each consecutive codon. For each tRNA, its amino acid is released and attached to a growing string of amino acids that will become a new protein.

Each tRNA used in the translation process loses its amino acid. Therefore the tRNA must be recharged with a new amino acid before it can be used again in the translation process. The recharging is accomplished by a type of protein known as aminoacyl-tRNA synthetase. Aminoacyl-tRNA synthetases attach an amino acid to the cognate tRNA. This amino acid-to-tRNA pairing, accomplished by the aminoacyl-tRNA synthetases, is the key to the genetic code. In other words, what we refer to as the genetic code is simply a representation of which amino acids are attached to which tRNA molecules. Therefore, it is the aminoacyl-tRNA synthetases that implement the genetic code.

This paper focuses on the well-accepted and widespread understanding that the genetic code is among the best evidence for common descent and evolution. How the code itself is thought to have evolved is relevant, but the main observations and logic that make the genetic code such powerful evidence for evolution lie elsewhere, in properties and aspects of the code and the greater biology.

According to evolution the species arose via gradual change and speciation. This implies that whereas similarities in distant species—such as the camera-like eye found in humans and squids—must have evolved independently, on the other hand similarities in more closely-related species—such as the eyes of different frog species—arose from a common ancestor. These two types of similarities are known as *analogies* and *homologies*, respectively, and it is the latter that are viewed as revealing evolution's common descent pattern and serving as powerful evidence for evolution. Homologies are interpreted as particularly powerful evidence for evolution when their structure appears to be arbitrary, functionless, harmful, or as Darwin often put it, trifling. (Darwin 1859, p. 426) In these cases, independent creation or design are thought to be much less likely, leaving common descent as by far the most likely explanation. (Sober 2008, pp. 264-314)

The genetic code, discovered in the mid twentieth century, was interpreted as the ultimate homology, and a powerful confirmation of evolutionary theory. This was primarily due to two reasons. First, the code was thought to be *mundane*, not having special capabilities or features compared to other possible codes. Second, the code—both its underlying molecular machinery, and the resulting codon-to-amino acid mapping—was thought to be conserved across different species and therefore *universal*.

This combination of mundane-ness and universality made the code powerful evidence for common descent and evolution. It was not as though the code had special properties making it the obvious choice for all the species. There was no functional reason for this particular, mundane, code. Furthermore, the code's universality made it the ultimate homology, spanning all life. But why? It did not seem to be chemically determined—the codons and the amino acids for which they code were disconnected. The code appeared to be an arbitrary assignment of codons to amino acids. And why should the same code be used in different species? Completely different codes in each species would work just fine. There was no reason for the code to be universal. And yet, the same code appeared in each species for no reason—a sure sign of common descent. Simply put, the two evidences were, first, there was no reason for this particular code, and second, there was no reason for the code to be universal.

While these arguments may appear to be powerful and compelling, the underlying details raise questions. This paper reconstructs and critically examines these arguments. This paper does not examine the question of how the code arose, except in passing where relevant to the topic of this paper: the argument that the genetic code is powerful evidence for evolution.

The genetic code as evidence for evolution

Not long after the painstaking elucidation of the genetic code in the 1960s, researchers arrived at two important conclusions: the genetic code appeared to be (i) not optimized or particularly special and (ii) essentially universal across the species. (Crick 1968) These would become important premises as the genetic code was not merely viewed as fitting well within the context of evolution but even more so the code was, as one paper put it, "widely considered the most persuasive evidence for UCA [universal common ancestry]." (Theobald 2011) So well accepted is the evidence that often it is merely mentioned in passing without justification, for it simply is a given that "The common ancestry of all extant cellular life is evidenced by the universal genetic code." (Moody 2024, p. 1654)

The conclusion that the genetic code is not special can be traced back at least to Francis Crick who argued that, while some structure in the code is obvious, "There is no reason to believe, however, that the present code is the best possible, and it could have easily reached its present form by a sequence of happy accidents." (Crick 1968, p. 377) So the genetic code was an "accident" (Crick 1981) that was "selected arbitrarily," as a widely used undergraduate molecular biology text later explained. (Alberts et al. 1994, p. 9) After all, "the code evolved early on in the history of life and … is then what Crick called a 'frozen accident.'" (Ridley 1993, p. 48) Hence the code was viewed as a contingency of history.

> This "genetic code" is arbitrary in the same sense that human language is arbitrary. For example, nothing forces English speakers to use the word "bird" to refer to animals with wings and feathers. It is merely a convention. Spanish speakers get the same idea across by saying "pajaro." (Rogers 2011, 30-31)

Coupled with this view that the code is essentially mundane, was the code's universality. As Crick explained, the code's universality immediately implied universal common ancestry: "To account for it being the same in all organisms one *must assume* that all life evolved from a single organism (more strictly, from a single closely interbreeding population)." (Crick 1968, p. 370, emphasis added) It was, as Ridley later explained, among "the best evidence that all life has a single common ancestor," for though the code is universal it "would not be if the species had independent origins." (Ridley 1993, p. 48-49) Just as the code's mundaneness showed there was no reason for this particular code, so too there was no reason for the code's universality.

> Different species could have different genetic codes just as easily as different human populations speak different languages. But they don't. Every living thing—from the smallest microorganism to the largest whale—makes protein using essentially the same arbitrary code. What sense does this make, unless all these forms of life evolved from a single ancestor? (Rogers 2011, 31)

Half a century after Crick, his early conclusion continues to obtain that "the close similarity of the genetic code in all known cellular life attests to universal [common ancestry] (Crick 1968)." (Baum et al. 2016, p.1354)

An Internal Contradiction

This high confidence that the genetic code is exemplary of the best evidence for common descent and evolution is, however, challenged by an internal contradiction. The code's universality implies its presence in the Last Universal Common Ancestor (LUCA) of all life. This suggests two inferences. First, the code must have arisen in a relatively short time period in early evolutionary history. Indeed, some argue that "the origin of the genetic code occurred simultaneously with the evolution of cellularity. That is to say … the origin of the genetic code is a very, very late event in the history of life on Earth." (Di Giulio 2024) Second, aside from very rare minor variants, the code shows a highly robust stability while, otherwise, evolutionary change crossed vast territories over billions of years. This second inference could be understood because any change to the code would disrupt the primary sequences of untold numbers of proteins. This "would be highly disadvantageous unless accompanied by many simultaneous mutations to correct the 'mistakes' produced by altering the code. This accounts for the fact that the code does not change." (Crick 1968, p. 370) As later accounts explained, "A mutation that caused the nucleotide sequence UUU to code for glycine instead of phenylalanine would have messed up all the species' proteins," (Futuyma 1982, p. 205) and once the code evolved "it would be strongly maintained. Any deviation from the code would be lethal." (Ridley 1993, 48)

These arguments that the code is not evolvable, however, have no principled reason for why they should not apply to the original evolution of the code. In other words, if the standard genetic code (SGC), once evolved, could not tolerate further evolutionary change, how could it tolerate the enormous amount of evolutionary change required to evolve the code in the first place? Hence, we must believe that the genetic code must have evolved, and yet it is not evolvable. Crick's solution to this internal contradiction was the "frozen accident" hypothesis. The code did not actually evolve in the first place, it fortuitously appeared. Natural selection played no role. But the code is highly complex. As one researcher explained, "The genetic code is this amazing thing … It's a mind-bogglingly complicated process." (Stolte 2024)

For example, the genetic code—which is used in protein synthesis—requires the pre-existence of a large number of proteins. So the code is required to construct proteins, but proteins are required to execute the code. How then could the code evolve without the prior existence of proteins, which in turn would require the prior existence of the code?

> what would be the selective force behind the evolution of the extremely complex translation system before there were functional proteins? And, of course, there could be no proteins without a sufficiently effective translation system. A variety of hypotheses have been proposed in attempts to break the circle but so far none of these seems to be sufficiently coherent or enjoys sufficient support to claim the status of a real theory. (Koonin and Novozhilov 2009, p. 108)

What is needed is an unlikely event, and one study estimated the probability of a replicator with protein synthesis type machinery evolving to be one in 10^{1018}. (Koonin 2007) Thus, Crick resolved the internal contradiction with his frozen accident hypothesis, but at the cost of requiring the surmounting of an enormous entropic barrier. It would be highly serendipitous for the genetic code somehow to arise accidentally.

After Crick, evolutionists proposed a multiplicity of hypotheses for the origin of the genetic code, including the stereochemical, coevolution, horizontal gene transfer, error minimization, ambiguity-reduction, four-column, RNA world, coding coenzyme handle, and gradual expansion hypotheses. There is a multiplicity of hypotheses because none of them are very convincing, realistic, detailed, or complete. Most of the hypotheses invoke non-empirical entities, such as biological worlds that have never been observed or mechanisms and pathways that are significantly different from what has been observed. For example, the error minimization hypothesis proposes that the genetic code evolved to minimize the impact of DNA mutations or translation errors—a process that has never been observed in nature. In this process the code would tend toward a design in which such mutations or errors would result in the substitution of a chemically similar amino acid (e.g., one hydrophobic amino acid for another hydrophobic). But why did such a process halt before reaching an optimum, and how could such a process occur if the code is unevolvable in the first place? In this sense, the hypothesis "takes vertical descent as its starting point, and thus to some extent assumes as given that which it sets out to explain." (Froese et al. 2018) Furthermore, the code has some structure not related to error minimization, and so the hypothesis seems at least to be incomplete.

Or again, the gradual expansion hypothesis proposes that the code evolved from a simpler code with fewer codons and amino acids, and expanded eventually to the SGC. Such a process has never been observed in nature and, again, there is the question of how such a process could occur if the code is unevolvable. This is highlighted by the "mystery" of how ancient genes could be "recoded" as the code evolved. (Wehbi et al. 2024)

Given the inadequacies of the various hypotheses that have been proposed, some researchers believe the code evolved via some combination of these hypotheses. (Koonin and Novozhilov 2017, p. 45) Such combinations, while adding theory complexity, do not provide compelling and detailed explanations for the code's evolution. Therefore it has seemed that after half a century we are no closer to a solution. (Froese et al. 2018) As one paper admitted, "To be honest, we are stuck ... there is an impasse." (Kun and Radványi 2018)

Reevaluation of the genetic code as evidence for evolution

Nothing here is predicated on the research into the origin of the genetic code to be at an impasse. But given the difficulties in generating a detailed, complete, realistic and compelling hypothesis for the genetic code evolution, this naturally raises the question of whether the code can truly provide such powerful evidence for common descent and evolution, as has been so often claimed. To examine that claim we must return to the two fundamental premises undergirding the argument. As discussed above, the genetic code is thought to be (i) not optimized or particularly special and (ii) essentially universal across the species. As we shall see, both premises are false.

The genetic code has structure that is immediately evident upon inspection. But a quantitative evaluation of the code's structure and uniqueness came in a 1998 paper entitled "The genetic code is one in a million." (Freeland and Hurst 1998) The title indicated how the SGC compares to randomly generated codes, in terms of its error minimization properties, as discussed above. Similarly, other studies discovered unique and special properties of the code. One found that the genetic code is a very rare code, even when compared to other codes which already have the error correcting capability. (Itzkovitz and Alon 2007) Another found that the code does not improve merely one function, but rather "a combination of several different functions simultaneously." (Bollenbach, Vetsigian and Kishony 2007, p. 403) As one paper concluded, the code's properties were "unexpected and still cry out for explanation." (Vetsigian, Woese and Goldenfeld 2006, p. 10696) And a later study upgraded the code from one in a million to one in 10^{20}. (Omachi et al. 2023) While we may still not fully understand the code's properties, it is clear that the premise that the code is not particularly special has failed. The code is far more unique and special than evolutionists had believed.

What about the second premise that the genetic code is essentially universal? First, recall that the genetic code concept entails both (i) the mapping of amino acids to codons and (ii) the molecular machinery that enact the code, in the cell. Evolutionists have both of these two phenomena in view when making the argument that the code's universality is powerful evidence for evolution. Indeed, under evolution, it is the universality of the molecular machinery that renders the universality of the mapping. The evidential claim would make no sense if the molecular machinery is not universal. Hence Niles Eldredge celebrated the "underlying chemical uniformity of life" as a severe test that evolution passed with flying colors. (Eldredge 1982, 41) And Michael Ruse concluded that the essential macromolecules of life help to make evolution beyond reasonable doubt. (Ruse 1986, 4) These essential macromolecules specifically included the aminoacyl-tRNA synthetases—the proteins that enact the genetic code by charging the tRNA molecules with their cognate amino acids.

> Indeed, the universality of the key components of the translation system *including a nearly complete set of aminoacyl-tRNA synthetases* among the extant cellular life forms strongly suggests that the main features of the translation system were fixed at a pre-LUCA stage of evolution. (Koonin and Novozhilov 2009, p. 106, emphasis added)
>
> It is common belief that all cellular life forms on earth have a common origin. This view is supported by the universality of the genetic code and the *universal conservation of multiple genes*, particularly those that encode key components of the translation system. (Koonin and Wolf 2010, emphasis added)

But this is false. While the premise requires that the cell's aminoacyl-tRNA synthetases are universal, they are not. These proteins differ dramatically across the species and present all manner of contradictions to the common descent pattern, "sometimes in spectacular ways." The code's machinery can only be said to be universal at the higher level of pathways and molecular entities: translation mechanisms using ribosomes, tRNAs and aminoacyl-tRNA synthetases are found throughout life. But the molecular sequences of the aminoacyl-tRNA synthetases are not universally conserved, and this has required an extensive use of horizontal gene transfer to attempt to rationalize in terms of common descent. But even with a liberal application of horizontal gene transfers, contradictions remain. For example, whereas "one would expect the horizontal gene transfer profiles of the 20 synthetases all to be similar in some general respect … this is clearly not the case: the horizontal gene transfer profiles of the different synthetases can be qualitatively different." For example, horizontal gene transfers are generally expected to be taxonomically local, but in order to fit the aminoacyl-tRNA synthetases to a common descent pattern, transfers are required that “are not predominantly local” as “transfers from one major taxon to another major taxon, appear to be unexpectedly prevalent.” Furthermore, it must be assumed that “the dynamic of horizontal gene transfer has not remained constant over the evolutionary course” and therefore “the nature of cells has changed over that course.” These are enormous implications for our understanding of biology and the history of life, all caused by the non universal patterns of the aminoacyl-tRNA synthetases. Hence, “the evolutionary process is not what it might seem” at the genome sequence level. Rather than a “more or less continuous progression leading from very simple primitive forms to the modern ones … evolution may proceed through a series of discontinuous stages.” (Woese et al. 2000)

Therefore, at the fundamental level of the aminoacyl-tRNA synthetase proteins, the assumption of a universal genetic code, meaning that the code—including its molecular machinery—is strongly conserved across all species, has failed. And the observed pattern is not explained by conventional evolutionary theory. Instead, radical, non parsimonious, deployment of horizontal gene transfer accompanied by dramatic changes to the history of life is required.

This undercuts the argument that the universality of the genetic code makes it a powerful, convincing evidence for common descent and evolution. That universality is required at the molecular level. Instead, we must believe that the aminoacyl-tRNA synthetases, as well as the course of life itself, underwent radical changes, contradicting the common descent pattern, while nonetheless preserving the code’s mapping of codons to amino acids. Clearly, this radical hypothesis is driven by and necessitated by the assumption of evolution, rather than the other way around. The argument that the code’s universality necessitates evolution loses its force.

But there is an even more serious failure here, for this evidential argument suffers from a fundamental scientific mistake. Recall that a key corollary—in the argument that the code’s universality implies common descent—is that there is no reason for the code’s universality. Therefore common descent is the only plausible explanation for the code’s universality. As Crick put it, “To account for [the genetic code] being the same in all organisms one must assume that all life evolved from a single organism …” (Crick 1968, p. 370) But of course that is not true. The code’s property of universality is crucial in biology, exclusive of common descent. The notion that there is *no reason* the genetic code should be conserved across the species fails badly given the importance of inter-specific transfer of genetic material, such as in the horizontal gene transfer process and RNA transfer mechanisms. (Valadi et al. 2007) It is astonishing that Crick’s false assertion not only has gone uncorrected, but that it has continued to be reaffirmed. Recall that Ridley claimed the code “would not be [universal] if the species had independent origins.” (Ridley 1993, p. 49) This refutation of the alternative is what gives the argument its power. (Sober 2008, pp. 264-314) But it contradicts the known biology. It would be safe to say that life as we know it would be impossible without the code’s universality. Of course the code must be universal. That is fundamental to biology, regardless of evolutionary theory. Yet in another example we saw Rogers perpetuate this argument:

> “Different species could have different genetic codes just as easily as different human populations speak different languages. But they don’t. Every living thing—from the smallest microorganism to the largest whale—makes protein using essentially the same arbitrary code.

> What sense does this make, unless all these forms of life evolved from a single ancestor?" (Rogers 2011, 31)

This is not merely false but a serious scientific mistake. It is crucial to the argument and persists, either explicitly or implicitly, in the literature, to argue that the genetic code proves common descent and evolution.

Reevaluation of the genetic code as a challenge for evolution

So far we have seen that textbook orthodoxy for roughly half a century—and tracing back to Crick—that the genetic code is powerful evidence for common descent and evolution, is repeatedly contradicted by the science. But the science does not merely reject the claim that the genetic code is evidence for evolution. In fact, much more so, the science presents the case that the genetic code is evidence against evolution. This challenge arises from the several evidences discussed above. But the above discussion was in the context of the genetic code as evidence for evolution. It is instructive to review the evidence without such a bias. What becomes apparent is that they pose a challenge for evolution.

First, the code's high complexity poses an enormous entropic barrier for a blind evolutionary process—by whatever mechanisms one chooses to assemble. One aspect of this complexity arises from the code's circularity. The code is at the heart of the cell's protein synthesis, yet many proteins are required to execute the code. In other words, proteins are first required in order to create proteins. Koonin's estimate of the probability of a replicator, with protein synthesis type machinery, evolving to be one in 10^{1018} (Koonin 2007) illustrates the magnitude of the problem.

Next, the very nature of the code is not as a bystander that can freely evolve without consequence. Instead, changes ripple through the cell's proteome, and so as evolutionists have argued (in order to rationalize the code's universality and hence robustness to change) "any deviation from the code would be lethal." (Ridley 1993, 48) This assertion that the code is unevolvable is not merely an argument from the nature of the code, it is also an argument to the code's universality. In other words, the code must be unevolvable because its universality requires the code to be essentially static, rejecting the vast majority of any change, while otherwise evolution traverses vast tracts of biological change, ranging from prokaryotes, archaea and eukaryotes, to multicellularity, fish, amphibians, reptiles, birds, mammals, and all their associated novel designs. Hence evolution was capable of generating expansive diversity, but not of altering the genetic code. It is one of the most striking examples in all of biology of an unevolvable design. And yet, of course the code must have evolved somehow.

In addition to the code being unevolvable, its universality also implies that the code was present in the LUCA, early in evolutionary history. This therefore imposes a relatively short time period for the code to arise. Furthermore, it was not merely any random code that evolved. Rather, the internal structure of the code was, surprisingly, found to be very special. Somehow this rare code, that otherwise is unevolvable, overcame an enormous entropic barrier to evolve in a relatively short time period, along with thousands of protein-coding genes that somehow recoded as the code evolved.

On top of all this, the code's fundamental molecular machinery, the aminoacyl-tRNA synthetases, were not conserved but rather must have undergone radical evolutionary changes unlike anything we normally observe. To explain the distribution of aminoacyl-tRNA synthetases across the species requires extensive and extreme levels of horizontal gene transfer, along with profound changes to life itself. All of this is not observed, and rather is inferred in order to fit the aminoacyl-tRNA synthetase sequences to common descent.

Evolutionary attempts, such as they are, to explain the origin of the genetic code and resolve these issues suffer from various combinations of incompleteness, serendipity, low probability, unobserved mechanisms, and loss of parsimony (violations of Occam's Razor). Rather than serving as powerful evidence for evolution as is routinely claimed, the genetic code contraindicates evolution.

Conclusions

Since its discovery in the mid twentieth century, the genetic code has been almost universally understood as powerful, convincing evidence for common descent and evolution. This claim is so well accepted it is often presented as a brute fact, without justification. This textbook orthodoxy, so embedded in the literature, is an unfortunate example of a science driven by doctrine rather than driven by data. We have allowed tradition and doctrine to infiltrate our science. Claims that are obviously false and dogmatic have gone uncorrected and reiterated in the literature.

This problem is all the more obvious by the challenges presented by empirical science. In other words, rather than powerful evidence for evolution, the genetic code presents an array of serious challenges to the theory. We have erred, and it is important for science in general, and evolution in particular, to follow and present the empirical evidence accurately and in an unbiased manner.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Artificial Intelligence: No artificial intelligence was used in the research or writing of this manuscript.

References

(Alberts et al. 1994) Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, J. Watson. 1994. *Molecular Biology of the Cell*. 3d ed. New York: Garland Publishing.

(Baum et al. 2016) Baum, David, Cécile Ané, Bret Larget, et al. 2016. "Statistical evidence for common ancestry: Application to primates." *Evolution* 70:1354–63.

(Bollenbach, Vetsigian and Kishony 2007) Bollenbach, T., K. Vetsigian, R. Kishony. 2007. "Evolution and multilevel optimization of the genetic code." *Genome Research* 17:401-404.

(Crick 1968) Crick, Francis. 1968. “The origin of the genetic code.” *J. Molecular Biology* 38:367-379.

(Crick 1981) Crick, Francis. 1981. *Life Itself: Its Origin and Nature*. New York: Simon and Schuster.

(Darwin 1859) Darwin, Charles. 1859. *The Origin of Species*. 1st ed. London: John Murray.

(Di Giulio 2024) Di Giulio, Massimo. 2024. “The time of appearance of the genetic code.” *BioSystems* 237:105159.

(Eldredge 1982) Eldredge, Niles. 1982. *The Monkey Business*. New York: Washington Square Press.

(Freeland and Hurst 1998) Freeland, S., L. Hurst. 1998. “The genetic code is one in a million.” *J. Molecular Evolution* 47:238-248.

(Froese et al. 2018) Froese, Tom, Jorge I. Campos, Kosuke Fujishima, Daisuke Kiga, Nathaniel Virgo. 2018. “Horizontal transfer of code fragments between protocells can explain the origins of the genetic code without vertical descent.” *Scientific Reports* 8:3532.

(Futuyma 1982) Futuyma, Douglas. 1982. *Science on Trial: The Case for Evolution*. New York. Pantheon Books.

(Itzkovitz and Alon 2007) Itzkovitz, S., U. Alon. 2007. “The genetic code is nearly optimal for allowing additional information within protein-coding sequences.” *Genome Research* 17:405-412.

(Koonin 2007) Koonin, Eugene. 2007. “The cosmological model of eternal inflation and the transition from chance to biological evolution in the history of life.” *Biology Direct* 2:15

(Koonin and Novozhilov 2009) Koonin, Eugene, A.S. Novozhilov. 2009. “Origin and evolution of the genetic code: the universal enigma.” *IUBMB Life*. 61:99-111.

(Koonin and Novozhilov 2017) Koonin, Eugene, A.S. Novozhilov. 2017. “Origin and Evolution of the Universal Genetic Code.” Annu Rev Genet. 51:45-62.

(Koonin and Wolf 2010) Koonin, Eugene, Y.I. Wolf. 2010. The common ancestry of life. *Biology Direct* 5:64.

(Kun and Radványi 2018) Kun, Á., Á. Radványi. 2018. “The evolution of the genetic code: Impasses and challenges.” *Biosystems* 4:217-225.

(Moody 2024) Moody, E.R.R., S. Álvarez-Carretero, T.A. Mahendrarajah, et al. 2024. “The nature of the last universal common ancestor and its impact on the early Earth system.” *Nature Ecology & Evolution* 8:1654-1666.

(Omachi et al. 2023) Omachi, Y., N. Saito, C. Furusawa. 2023. “Rare-event sampling analysis uncovers the fitness landscape of the genetic code.” *PLoS Comput Biol* 19:e1011034.

(Ridley 1993) Ridley, Mark. 1993. *Evolution*. Boston: Blackwell Scientific.

(Rogers 2011) Rogers, Alan. 2011. *The Evidence For Evolution*. Chicago: University of Chicago Press.

(Ruse 1986) Ruse, Michael. 1986. *Taking Darwin Seriously*. New York: Basil Blackwell.

(Stolte 2024) Stolte, Daniel. 2024. “Study sheds light on origin of genetic code.” *University of Arizona*, Dec. 12, 2024.

(Sober 2008) Sober, Elliott. 2008. *Evidence and Evolution: The Logic Behind the Science*. Cambridge: Cambridge University Press.

(Theobald 2011) Theobald D.L. 2011. “On universal common ancestry, sequence similarity, and phylogenetic structure: the sins of P-values and the virtues of Bayesian evidence.” *Biol Direct*. 6:60.

(Valadi et al. 2007) Valadi, Hadi, Karin Ekström, Apostolos Bossios, et al. 2007. “Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells.” *Nat Cell Biol* 9:654-9.

(Vetsigian, Woese and Goldenfeld 2006) Vetsigian, K., C. Woese, N. Goldenfeld. 2006. “Collective evolution and the genetic code.” *Proceedings of the National Academy of Sciences* 103:10696-10701.

(Wehbi et al. 2024) Wehbi, Sawsan, Andrew Wheeler, Benoit Morel, et al. 2024. “Order of amino acid recruitment into the genetic code resolved by last universal common ancestor’s protein domains.” Proc. Natl. Acad. Sci. U.S.A. 121 (52) e2410311121.

(Woese et al. 2000) Woese C.R., G.J. Olsen, M. Ibba, D. Söll. 2000. “Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process.” *Microbiol. Mol. Biol. Rev*. 64:202–236.