OPEN ACCESS

BIO-complexity

Research Article

On the Origin of the Codes: The Character and Distribution of Variant Genetic Codes is Better Explained by Common Design than Evolutionary Theory

Winston Ewert*
Biologic Institute, Redmond, Washington, USA

Abstract

The near universality of the genetic code is frequently cited as evidence for universal common ancestry. On the other hand, critics of universal common ancestry frequently point to exceptions to the universal code as evidence against it. However, there has never been a comprehensive investigation into the character and distribution of variant genetic codes and their implications for the debate over universal common ancestry. This paper develops a framework for understanding codes within a common design framework, based crucially on the premise that some genetic code variants are designed and others are the result of mutations to translation machinery. We found that these two sources of variant codes can be distinguished by considering organismal lifestyle, taxonomic rank, evolutionary feasibility, codon rarity and complexity of distribution. These different approaches to distinguishing the codes give highly correlated results, demonstrating impressive explanatory power for our framework. In contrast, we find that evolutionary theory has difficulty explaining the character and distribution of variant genetic codes.

Cite As: Ewert W (2024) On the Origin of the Codes: The Character and Distribution of Variant Genetic Codes is Better Explained by Common Design than Evolutionary Theory. BIO-Complexity 2024 (1):1-25. **doi:**10.5048/BIO-C.2024.1.

Editor: Robert J. Marks II

Received: September 19, 2023; **Accepted:** January 9, 2024; **Published:** March 2, 2024

Notes: A *Critique* of this paper, when available, will be assigned **doi:**10.5048/BIO-C.2024.1.c.

*Email: evoinfo@winstonewert.com

Most genomes use the canonical, or universal, genetic code. That is, they use the same mapping between codons and amino acids or stop codons. (There are differing start codons or non-standard amino acids like selenocysteine or pyrrolysine; however, for our purposes, we will consider only the mapping between codons and standard amino acids or stop codons) So then, most genomes decode their codons according to the canonical genetic code.

Those genomes that use a non-canonical or variant genetic code use codes highly similar to the canonical code, differing in only a few codon mappings. This near-universality of the genetic code is frequently invoked as crucial evidence for common descent [1–7]. Critics of evolutionary theory often focus on the variant codes, arguing that they undermine evolutionary theory [8, 9]. However, there has been no comprehensive attempt to investigate the character and distribution of variant genetic codes and what they mean for the origin of the codes.

Upon first investigation, evolutionary theory appears to have a compelling account of the character and distribution of variant codes. Evolutionary theory suggests that if genetic code evolution is possible it should be very rare. This would explain why most genomes follow the standard code and why the exceptions only vary in a few codons. It would also explain the following details about the variant codes. Most variations are found in mitochondria, whose very small genomes would make code evolution easier. Many variations are also found in highly reduced genomes, such as those of endosymbiotic bacteria. No variations are found in the nuclear genomes of complex multicellular organisms like plants and animals. The distribution of many codes can be easily explained by identifying certain points on the tree of life where codons were reassigned and then inherited by all of their descendants.

Evolutionary theories of code evolution have been proposed, including the codon capture model [10], the

ambiguous intermediate model [11] or the loss-driven codon reassignment model [12]. Critics have pointed out that these models require highly improbable events such as the fixation of highly deleterious mutations or the total disappearance of a codon. They argue that this undermines evolutionary theory's ability to explain the existence of variant codes. However, our primary purpose here is not to critique evolutionary theory's ability to explain code evolution but rather to develop an alternative understanding of the character and distribution of variant codes. Towards that end, Section 1 will develop a common design framework for variant codes. Section 2 will discuss a survey of all known variant genetic codes in bacterial, nuclear, and mitochondrial genomes. Section 3 will flesh out the common design framework, demonstrating its explanatory power for the distribution and character of variant genetic codes by demonstrating five different, highly correlated criteria for distinguishing designed and evolved codes. Section 4 will look at evolutionary theory's interpretation of this data and demonstrate that, despite initial appearances, it has poor explanatory power for the character and distribution of variant genetic codes.

1. FRAMEWORK

In this work, we put forward a framework that seeks to explain the character and distribution of variant genetic codes within a common design framework. Common design is the idea that commonalities in living things are due to having been designed by a common designer rather than having descended from a common ancestor. It is important to note that advocates of common design do not hold to species fixity: the independent design of every extant species. Instead, they accept a limited form of common ancestry where extant species thought to be closely related do, in fact, descend from a common ancestor. Limited common ancestry is often described as an orchard model. In contrast to the single tree of life under standard evolutionary theory, numerous independent trees connect limited groups of species.

Our previous work has begun to develop a common design model, in particular accounting for the existence of higher taxa [13, 14]. That work explains the higher taxa in terms of modules connected by a dependency graph. We will not go into the details of that model here, except to note that it attempts to explain the apparent nested hierarchy pattern found in biological life. Insofar as variant genetic codes follow the nested hierarchy, we appeal to the dependency graph model as an explanation. Nevertheless, familiarity with the dependency graph model should not be required for readers of this paper unless they are specifically concerned about the hierarchical pattern.

The first of three tenets of our framework is that the canonical genetic code has been well optimized and is thus an ideal choice for most genomes. Research into the optimized nature of the genetic code has a long history [15–17]. It is important to appreciate that the genetic code is not merely optimized according to one criterion but according to many different criteria. For example, similar codons are assigned to similar amino acids, minimizing damage caused by misreadings or mutations [18]; the number of stop codons is optimized to trade off between read-through errors and premature terminations [19]; and the identity of the stop codons is optimized to increase the probability of appearing in frame-shifted gene sequences [20]. Various other optimized criteria have been observed, and more probably remain to be discovered.

It is important to appreciate the role of design trade-offs. There are numerous different criteria by which the quality of the genetic code could be measured. By necessity, improving the genetic code according to one criterion will require degrading it according to other criteria. A designer must identify the best trade-offs to select the ideal genetic code. Life is diverse, but all lifeforms primarily use the same twenty amino acids, many of the same proteins and similar translation machinery. As such, for most genomes, the trade-offs between different codes are similar. Thus, the canonical code is usually an ideal choice.

The second of the three tenets of our framework is that a slightly tweaked version of the canonical code is better for some organisms. Recall that the canonical genetic code reflects a particular set of trade-offs that make sense for most genomes. Nevertheless, some genomes will be different in some way, altering the ideal trade-off. Utilizing a different genetic code with a somewhat different set of trade-offs in those genomes would make sense. Furthermore, there are likely many codes similar to the canonical genetic code with slightly different trade-offs. As such, the designer can improve some genomes by adopting a minor variation on the canonical genetic code for those genomes.

The third of the three tenets is that mutations in some organisms have damaged the translation machinery, causing them to misinterpret the code they were designed to use. Consequently, these organisms now translate the genetic code differently and thus operate on a variant genetic code. Such codes were not designed but instead evolved. This tenet is expected to be controversial among critics of evolutionary theory, who have tended to defend the impossibility of any sort of evolutionary code change. As will be explained in Section 3.3, we have laboratory evidence for these sorts of changes being possible. Additionally, we will see that, in some cases, these variant codes are better understood as flawed interpretations of the standard code than distinct variant codes.

In summary, three tenets define our basic framework. Firstly, the canonical genetic code is well-engineered and

suitable for most genomes. Secondly, variations on the canonical genetic code are sometimes a better choice for specific genomes. Thirdly, some extant species have acquired new genetic codes through mutations to their translation machinery. Together, these tenets provide a framework for explaining the distribution and character of the genetic codes within common design.

2. CODE VARIANTS

There is no comprehensive database or review paper collecting all known variant genetic codes. The NCBI taxonomy lists the genetic codes for all species in the taxonomy [21]; however, these listings often do not reflect the state of the literature. In order to construct a comprehensive collection of variant codes, we collected papers either publishing newly discovered codes or reviewing previously published code discoveries. From this, we constructed a database consisting of all genomes for which the genetic code has been reported in the literature. We used a simple parsimony rule to infer the codes used by organisms for which no code has been reported. Generally, we have followed the NCBI hierarchy to provide a phylogeny, but where the papers that reported a code gave a different phylogeny, we have followed the phylogeny presented in those papers.

In addition to recognized variant codes, many eukaryotes utilize extensive stop codon readthrough [22–24]. Despite sometimes decoding what is usually a stop codon as an amino acid, these organisms are generally still considered to follow the standard code. However, this variegated interpretation is also found in some organisms that follow "stopless" codes, which reassign all three stop codons. In such organisms, at least some canonical stop codons still function as stop codons but only some of the time. The difference between the stopless codes and the standard code with extensive readthrough is one of degree rather than kind. A comprehensive study of genetic codes would ideally include the consideration of species with extensive readthrough. However, we will have to leave that to future research.

Eukaryotic organisms have both a nuclear genome and organelle genomes. Mitochondrial genomes are particularly important as they exhibit many well-studied variant codes. Other variant codes, such as those in plastids, are also known, but only a few examples have so far been discovered [25]. Due to a lack of data on those other codes, this paper focuses on bacterial, eukaryotic nuclear and eukaryotic mitochondrial genomes and will disregard the variant codes found in other organelles.

In this paper, we use the IUPAC notation as shown in Table 1. Table 2 lists the variant codes collected through a literature review.

Table 1: IUPAC base notation used throughout this paper.

A	Adenine
C	Cytosine
G	Guanine
U	Uracil
Y	Cytosine or Uracil (C or U)
R	Adenine or Guanine (A or G)
N	Any (A or G or C or U)

3. EXPLANATORY POWER

3.1 Introduction

This section will develop five different criteria proposed to distinguish evolved and designed codes within our common design framework. Though approximate, these criteria should each give us some idea of which codes are designed and which are evolved. Crucially, the distinction we are trying to make is to determine which codes are likely to be evolved or designed within the context of our proposed framework, not whether or not it is feasible to explain them evolutionarily. We will find that despite taking different approaches, these different criteria correlate well with each other.

Sections 3.2–3.6 develop the five criteria. Firstly, we expect that evolved codes should be found in taxonomic ranks below the level of family (the "Low Level Taxon" criterion), whereas designed codes will be found at or above the level of family (see Section 3.2). Secondly, we expect that any evolved code should be explicable in terms of some simple mutation to the translation machinery of the cell (the "Simple Mutation" criterion) (see Section 3.3). Thirdly, we expect that evolved codes are only found in endosymbiont genomes (the "Endosymbiont" criterion) (see Section 3.4). Fourthly, we expect that evolved codes involve a sufficiently small number of codons such that the change would not be too deleterious (the "Low Codon Usage" criterion) (see Section 3.5). Fifthly, we expect that evolved codes follow a simple phylogenetic distribution (the "Simple Distribution" criterion) (see Section 3.6).

Section 3.7 will show that these different criteria are well correlated. Section 3.8 will look at those cases that fit less well, showing how they make sense within the framework. Section 3.9 will briefly look at possible design trade-offs that might explain why different genomes might benefit from different codes.

3.2 Low Level Taxon

There is a wide range of taxonomic ranks for genetic code variations. Some are restricted to a particular species, while others are characteristic of high-level taxa up to the level of an entire kingdom. This is important to

Table 2a: Known variant codes for bacteria and protists (nuclear)

Index	Reassignment	Taxon name (level)	Crit. 1: Low level taxon	Crit. 2: Simple mutation	Crit. 3: Endosymbiont	Crit. 4: Low codon usage	Crit. 5: Simple distrib.	Evolved
bacteria								
1	UGA (Stop → Gly)	Gracilibacteria [26] (phylum) Absconditabacteria [26] (phylum)	No	No	No	Yes	No	No
2	UGA (Stop → Trp)	*Stammera capleta* [27] (species)	Yes	Yes	Yes	Yes	Yes	Yes
3	UGA (Stop → Trp)	*Hodgkinia cicadicola* [28] (species)	Yes	Yes	Yes	Yes	Yes	Yes
4	UGA (Stop → Trp)	*Nasuia deltocephalinicola* [29] (species)	Yes	Yes	Yes	Yes	Yes	Yes
5	UGA (Stop → Trp)	*Zinderia quadrilineatus* [30] (species)	Yes	Yes	Yes	Yes	Yes	Yes
6	UGA (Stop → Trp)	Mycoplasmatales [26] (order) Entomoplasmatales [26] (order)	No	Yes	Yes	Yes	Yes	No
7	CGR (Arg → Trp)	Absconditabacteria [26] (phylum)	No	No	No	No	No	No
8	CGG (Arg → Trp)	*Anaerococcus* [26] (genus)	Yes	Yes	Yes	Yes	Yes	Yes
9	CGG (Arg → Gln)	*Peptacetobacter* [26] (genus)	Yes	Yes	Yes	Yes	Yes	Yes
10	CGG (Arg → Trp)	Bacillales sp. [26] (related species)	Yes	Yes	Yes	Yes	Yes	Yes
11	AGG (Arg → Met)	Bacillales sp. [26] (related species)	Yes	Yes	Yes	Yes	Yes	Yes
protist (nuclear)								
12	UGA (Stop → Trp)	Blastocrithidia sp. [31] (related species)	Yes	Yes	Yes	Yes	Yes	Yes
	UAR (Stop → Glu)		Yes	Yes	Yes	Yes	Yes	Yes
13	UGA (Stop → Trp)	*Amoebophrya sp.** [32] (species)		Yes	Yes	Yes	Yes	Yes
	UAR (Stop → Gln)			Yes	Yes	Yes	Yes	Yes
14	UAR (Stop → Gln)	Hexamitinae [33] (subfamily)	Yes	Yes	Yes	Yes	Yes	Yes
15	UAR (Stop → Gln)	*Streblomastix* [34, 35] (genus)	Yes	Yes	Yes	Yes	Yes	Yes
16	UAR (Stop → Gln)	*Amoeboaphelidium protococcarum** [36] (species)		Yes	Yes	Yes	Yes	Yes
17	UAG (Stop → Gln)	*Iotanema spirale** [37] (species)		Yes	Yes	Yes	Yes	Yes
18	UAG (Stop → Leu)	*Rhizaria sp. exLh** [37] (species)		Yes	Yes	Yes	Yes	Yes

* The distribution of this variant is highly unclear due to limited reports on related genomes.

Table 2b: Known variant codes for ciliates (nuclear), algae (nuclear) and fungi (nuclear)

Index	Reassignment	Taxon name (level)	Crit. 1: Low level taxon	Crit. 2: Simple mutation	Crit. 3: Endosymbiont	Crit. 4: Low codon usage	Crit. 5: Simple distrib.	Evolved
ciliate (nuclear)								
19	UGA (Stop → Trp)	Plagiopylea* [38] (order)		No	No	No	No	No
	UAR (Stop → Gln)			No	No	No	No	No
20	UGA (Stop → Trp)	Blepharismidae [35, 39–42] (family)	No	No	No	No	No	No
21	UAA (Stop → Lys)	Oligohymenophorea sp. PL0344* [38] (order)		No	No	No	No	No
	UAG (Stop → Glu)			No	No	No	No	No
22	UAR (Stop → Gln)	Nassulidae [38] (family)	No	No	No	No	Yes	No
23	UAR (Stop → Gln)	Karyorelictea [39] (class)	No	No	No	No	No	No
24	UAR (Stop → Gln)	Condylostomatidae [39–42] (family)	No	No	No	No	No	No
25	UAR (Stop → Gln)	Cyrtolophosididae [41, 42] (family)	No	No	No	No	Yes	No
26	UAR (Stop → Gln)	Oligohymenophorea [35, 40–43] (class)	No	No	No	No	No	No
27	UAR (Stop → Gln)	Choreotrichia [40–42] (subclass) Stichotrichia [35, 40–42] (subclass) Oligotrichia [40–42] (order)	No	No	No	No	No	No
28	UAR (Stop → Glu)	Sessilida [40–42] (order)	No	No	No	No	No	No
29	UAR (Stop → Tyr)	Cyclotrichida [40–42] (order)	No	No	No	No	Yes	No
30	UGA (Stop → Cys)	Euplotida [35, 40–42] (order)	No	No	No	No	No	No
algae (nuclear)								
31	UAR (Stop → Gln)	Trentepohliales* [44] (order)		No	No	No	No	No
32	UAR (Stop → Gln)	Dasycladales [44] (order)	No	No	No	No	No	No
33	UAR (Stop → Gln)	Cladophorales [44] (order) *Blastophysa* [44] (genus)	No	No	No	No	No	No
fungi (nuclear)								
34	CUG (Leu → Ser)	*Metschnikowiaceae* [45] (family) *Debaryomycetaceae* [45] (family)	No	No	No	No	No	No
35	CUG (Leu → Ser)	*Saccharomycopsidaceae* [45] (family) *Ascoideaceae* [45] (family)	No	No	No	No	No	No
36	CUG (Leu → Ala)	*Pachysolenaceae [46]* [45**?**] (family)	No	No	No	No	No	No

* The distribution of this variant is highly unclear due to limited reports on related genomes.

Table 2c: Known variant codes for protists (mitochondrial) and algae (mitochondrial)

Index	Reassignment	Taxon name (level)	Crit. 1: Low level taxon	Crit. 2: Simple mutation	Crit. 3: Endosymbiont	Crit. 4: Low codon usage	Crit. 5: Simple distrib.	Evolved
protist (mitochondrial)								
37	UGA (Stop → Trp)	*Acanthamoeba castellanii** [43] (species)		Yes	Yes	Yes	Yes	
38	UGA (Stop → Trp)	*Amoebidium parasiticum** [43] (species)		Yes	Yes	Yes	Yes	
39	UGA (Stop → Trp)	*Cafeteria roenbergensis** [43] (species)		Yes	Yes	Yes	Yes	
40	UGA (Stop → Trp)	Prymnesiophyceae [43] (subclass)	No	Yes	Yes	Yes	No	
41	UGA (Stop → Trp)	Thalassiosirales [35, 43] (order)	No	Yes	Yes	Yes	No	
42	UAR (Stop → Tyr)	LAB14 [48] (uncultivated)		Yes	No	Yes	No	
	AGR (Arg → Stop)			No	No	Yes	No	
43	UAG (Stop → Tyr)	*Aplanochytrium* [48] (genus)	Yes	Yes	Yes	Yes	Yes	
44	UUA (Leu → Stop)	Labyrinthulomycetes [48] (class)	No	No	Yes	Yes	No	
	AUA (Ile → Met)		No	No	Yes	Yes	No	
45	AGR (Arg → Ser)	MAST8b [48] (uncultivated)		No	No	Yes	No	
	AUA (Ile → Met)			No	Yes	Yes	No	
algae (mitochondrial)								
46	UGA (Stop → Trp)	*Chondrus crispus** [35, 43] (species)		Yes	Yes	Yes	Yes	
47	UGA (Stop → Trp)	*Porphyra purpurea** [43] (species)		Yes	Yes	Yes	Yes	
48	UGA (Stop → Trp)	*Pedinomonas minor** [43, 49] (species)		Yes	Yes	Yes	Yes	
49	UGA (Stop → Trp)	Pycnococcaceae* [49] (family)		Yes	Yes	Yes	No	
	UUR (Leu → Stop)			No	No	Yes	No	
	AUA (Ile → Met)			No	Yes	Yes	No	
50	UGA (Stop → Trp)	Oligohymenophorea [43] (class)	No	Yes	Yes	Yes	No	
51	UAG (Stop → Ala)	Hydrodictyaceae [50] (family) Neochloridaceae [50] (family)	No	No	Yes	No	No	
52	UAG (Stop → Leu)	Scenedesmaceae [50] (family)	No	Yes	Yes	No	No	
53	UCR† (Ser → Stop)	Bracteacoccaceae [50] (family) Selenastraceae [50] (family) Scenedesmaceae [50] (family) Hydrodictyaceae [50] (family) Neochloridaceae [50] (family) Chromochloridaceae* [50] (family)	No	No	Yes	No	No	
54	CGG (Arg → Leu)	Sphaeropleaceae* [48] (family)		No	Yes	No	No	
55	CGG (Arg → Leu)	Chromochloridaceae* [48] (family)		No	Yes	No	No	
	AGR† (Arg → Met)			No	Yes	No	No	

* The distribution of this variant is highly unclear due to limited reports on related genomes.
† Some codons covered by the reassignment are unused.

Table 2d: Known variant codes for fungi (mitochondrial) and metazoans (mitochondrial)

Index	Reassignment	Taxon name (level)	Crit. 1: Low level taxon	Crit. 2: Simple mutation	Crit. 3: Endosymbiont	Crit. 4: Low codon usage	Crit. 5: Simple distrib.	Evolved
fungi (mitochondrial)								
56	UGA (Stop → Trp)	Dikarya [43, 51] (subkingdom)	No	Yes	Yes	Yes	No	
57	UGA (Stop → Trp)	Chytridiomycetes [43, 52] (class)	No	Yes	Yes	Yes	No	
58	AUA (Ile → Met)	*Eremothecium* [45, 51] (genus)	Yes	No	Yes	No	No	
	CUN (Leu → Ala)		Yes	No	No	No	No	
59	AUA (Ile → Met)	*Saccharomyces* [45, 51] (genus)	Yes	No	Yes	No	No	
		Nakaseomyces [45, 51] (genus)						
60	CUN (Leu → Thr)	*Kluyveromyces* [51] (genus)	Yes	No	No	No	No	
		Lachancea [51] (genus)						
		Saccharomyces [51] (genus)						
		Nakaseomyces [51] (genus)						
animal (mitochondrial)								
61	UGA (Stop → Trp)	Metazoa [43, 51, 53–57] (kingdom)	No	Yes	Yes	Yes	No	
		*Choanoflagellata** [43] (genus)						
62	UAG (Stop → Tyr)	*Clathrina clathrus*† [58] (species)	Yes	Yes	Yes	Yes	Yes	
63	UAG (Stop → Tyr)	*Leucetta chagosensis*† [58] (species)	Yes	Yes	Yes	Yes	Yes	
64	UAA (Stop → Tyr)	*Radopholus similis* [59] (species)	Yes	Yes	Yes	Yes	Yes	
		Radopholus arabocoffeae [59] (species)						
65	UAA (Stop → Tyr)	*Cephalodiscus hodgsoni** [54] (species)		Yes	Yes	Yes	Yes	
66	AGR (Arg → Stop)	Vertebrata [51, 54, 60] (subphylum)	No	No	No	No	No	
67	AGR (Arg → Ser)	Bilateria [43, 51, 53, 54] (group of phyla)	No	No	No	No	No	
68	AGR (Arg → Ser)	Calcarea [58] (class)	No	No	No	No	No	
69	AGR (Arg → Ser)	Hexactinellida [61] (class)	No	No	No	No	No	
70	AGR (Arg → Gly)	Tunicata [43, 51, 54] (subphylum)	No	No	No	No	No	
71	CGN (Arg → Gly)	Hexasterophora [58] (subclass)	No	No	No	Yes	No	
72	AUA (Ile → Met)	Catenulida [53] (class)	No	No	Yes	No	No	
73	AUA (Ile → Met)	Annelida [43, 51] (phylum)	No	No	Yes	No	No	
		Mollusca [43, 51] (phylum)						
		Brachiopoda [43, 51] (phylum)						
74	AUA (Ile → Met)	Ecdysozoa [43, 51, 59] (group of phyla)	No	No	Yes	No	No	
75	AUA (Ile → Met)	Chordata [33, 43, 51, 54, 60] (phylum)	No	No	Yes	No	No	
76	AAA (Lys → Asn)	Trematoda [43, 51, 53] (class)	No	No	Yes	No	No	
		Cestoda [51, 53] (class)						
		Monogenea [51, 53] (class)						
		Rhabditophora [53] (class)						
77	AAA (Lys → Asn)	Echinodermata [43, 51, 54] (phylum)	No	No	Yes	No	No	
78	AGG (Arg → Lys)	Various Arthropods [51, 56, 57] (various ranks)		No	Yes	No	No	
79	AGG (Arg → Lys)	Pterobranchia [54, 55] (class)	No	No	Yes	Yes	No	

* The distribution of this variant is highly unclear due to limited reports on related genomes. †The source suggests that UAG (Stop → Tyr) is found throughout the subclass on the basis of two distantly related species. However, inspection of GenBank reveals other mitochondrial genomes in the subclass make little use of UAG, suggesting it generally remains a stop codon.

any framework accepting limited common ancestry since these frameworks hold that many low-level taxonomic groups evolved from a common ancestor. As such, variant codes within these low-level taxonomic groups must have evolved rather than been designed. This gives us our first and most inflexible criterion for differentiating designed and evolved codes.

At what taxonomic level does descent apply? We propose that this is usually about the taxonomic rank of family. This aligns with a common suggestion in baraminology [62], the basic types as defined by Siegfried Schere [63, 64], the family line proposed by Behe [65] and what Ernst Mayr describes as "what a lay person would designate a 'kind of animal'" [66]. Consequently, we think that codes that are characteristic of families or higher taxa are usually designed. However, codes restricted to taxa below families (such as genera or species) are usually evolved. Thus, we can use the taxonomic rank as a first approximation of whether or not a code was designed or evolved. Nevertheless, this criterion is only an approximation; we expect some exceptions to exist.

There are a number of codes marked with an asterisk in Table 2 to indicate uncertainty in how widely distributed the reassignment is within that taxon. For particular high-level taxa, only members of a single genus, often a single species, have been studied to determine their genetic code. Discovered reassignments may be restricted to that single species or may, in fact, exist across that high-level taxon. Thus, for such cases, we have not included either *Yes* or *No* for the "Low Level Taxon" criterion in Table 2, since its status is unknown. For example, some members of the genus *Ameobophyra* reassign all of the stop codons; however, we found no reports of the genetic codes used by any other members of its class, Aphelidea, and so its distribution is uncertain. We have listed the code reassignment as applying to the taxa that make sense within our framework. We await further research to clarify the distribution of these variant codes.

3.3 Simple Mutation

Within our framework, we expect that some variant genetic codes have arisen by mutations to the translation machinery. This requires the existence of mutations that could explain the reassignments of evolved codes. Mutations that alter a cell's interpretation of its genetic code have been studied since the genetic code was first being deciphered [67–73]. However, in our framework other codes are designed. It is plausible that many designed reassignments would require intricate changes to translation machinery, which would be beyond the reach of mutation and selection.

There are three areas of research that give us particular insight into how a genome's interpretation of the genetic code might be altered by mutation. Firstly, stop-suppressor mutations can occur [69]. These were originally discovered due to their ability to allow cells to successfully translate a gene despite it containing a premature stop codon, hence the term *stop-suppressor*. Further research revealed that these mutations modified copies of tRNAs in the genome so that their anticodons matched the premature stop codons. The codons were still interpreted as stop codons most of the time, allowing these cells to continue functioning normally. However, the relevant codon was sometimes interpreted as a sense codon, allowing the gene to still partially function.

Secondly, other researchers have inserted tRNAs with modified anticodons targeting existing sense codons [70–72]. These modified tRNAs are ineffective compared to the native tRNAs targeting those codons but, nevertheless, successfully induce limited amounts of gene production with the codon translated according to a reassigned meaning.

Thirdly, researchers have disabled release factors, notably in an experiment in *E. coli* [73]. This prevented the cell from recognizing one of the stop codons, in this case, UAG. Despite the loss of the UAG stop codon, the cell was able to survive and reproduce. It was found that the absent stop codon was transcribed using near-cognate tRNAs as tyrosine, glutamine and tryptophan, allowing translation to proceed. The UAG codon was only used to terminate about 7% of the genes in *E. coli*. Of those genes, only seven were essential genes. All seven of those genes had another stop codon within 60 codons of the now non-functional UAG stop codon. As such, the loss of the UAG stop codon was not fatal.

These lines of research demonstrate that it is possible to alter a genome's interpretation of the genetic code through simple mutations. In particular, we see alterations by either mutating the anticodon of a tRNA or by disabling a release factor. However, this research also tells us which genetic code modifications are possible using a single, simple mutation and which are much harder to accomplish. In particular, there are three features required to facilitate the easiest case of a genetic code modification: near cognate tRNAs, redundant tRNAs and limited tRNA competition.

The first feature requires that a codon can only be reassigned to an amino acid with near-cognate tRNAs. That is, a codon can only be assigned to a meaning that is already assigned to a similar codon. By similar, we mean near-cognate, which means that it agrees in two of the three codon positions. Thus, for example, the UAA codon could be reassigned to glutamine because CAA, near-cognate to UAA, is already assigned to glutamine. However, UAA could not be reassigned to alanine because none of the codons assigned to alanine (GCU, GCC, GCA, and GCG) are near-cognate to UAA.

Why must reassignment have this feature? Reassigning codons to noncognate tRNAs is possible but would

require multiple changes. A single change to an existing tRNA can only modify one position of the anticodon. As a consequence, the newly modified tRNA must be near-cognate to the original tRNA. Any newly matched codons must have been near-cognate to the tRNA already. Likewise, if a tRNA or release factor is disabled or impaired, experimental research indicates that the codons end up being matched with near-cognate tRNAS, especially those that differ only in the wobble position [73]. Consequently, in any simple case of a mutation to the translation machinery, any reassignment will be to near-cognate tRNAs.

The second feature is a requirement for the modified tRNA to be redundant. If there is only a single tRNA translating a particular codon, modifying that tRNA to instead match other codons will leave the original codon untranslatable. This would be fatal in most cases but would not occur in most bacterial or nuclear genomes—these typically have redundant tRNAs. However, mitochondrial genomes typically contain a minimal number of tRNAs and therefore lack redundant tRNAs. This makes it difficult to modify tRNAs in mitochondrial genomes. It would be possible to overcome this difficulty if a tRNA were duplicated and then modified. Nevertheless, that is a more complicated scenario, beyond the reach of a single, simple mutation.

The redundancy requirement can be also be avoided if the tRNA can be modified so that it matches new codons while continuing to match the original codons. This is most readily possible if the new and old codons differ only in the wobble position. In the case of mitochondria, a single tRNA is often used to match all four codons that only differ in the wobble position. As such, the lack of redundant tRNAs does not provide a strong barrier to reassignments within a box, i.e. four codons that share all about the wobble position.

However, at least AGG (Arg → Lys) [55, 57] and AGR (Arg → Gly) [74] appear to be matched by tRNAs that match codons that differ in more than just the wobble position. The mechanics of AGG (Arg → Lys) are not clear. However, in the case of AGR (Arg → Gly) this is accomplished by RNA editing introducing a non-standard base, not a simple mutation [74]. We conclude that these kinds of reassignments require something more than the single, simple mutations considered here.

The third feature is the requirement for limited competition. Typically, if a tRNA is modified to match new codons, other translation machinery will already match those codons. Furthermore, research has shown that additional modifications to the tRNA are necessary in order to fine-tune it to match its new codons [70]. As such, a tRNA that has merely had its anticodon changed will be at a disadvantage relative to the existing tRNAs.

However, we can identify two scenarios where that competition will be at a minimum. The first is where a stop codon ceases to be recognized by a release factor. The most plausible outcome is that the codon then becomes translated by the most closely related tRNA. In the case of UGA, this suggests that it should be translated by the tRNA for the UGG codon and thus be decoded as tryptophan. In the cases of UAA and UAG, these should be translated by the tRNA for UAY and thus be translated as tyrosine. As such, reassignments such as UAA (Stop → Tyr), UAG (Stop → Tyr) and UGA (Stop → Tyr) appear plausible.

The second case that minimizes competition is a tRNA mutation that targets codons with a guanine in the wobble position. Due to wobble base pairing, most nucleotides in the wobble position bind to multiple nucleotides on the codon. The one exception is cytosine, which only binds to guanine. This makes it easier for a tRNA to specifically target a codon with a guanine in the wobble position than other codons. Additionally, these same codons are often matched using uracil, which, while it can match guanine via wobble base pairing, is at a disadvantage relative to cytosine. This gives an advantage to a new tRNA targeting the guanine codon using a cytosine nucleotide in the anticodon. As such, it is possible that a mutated tRNA targeting a codon with guanine in the wobble position could compete well enough to reassign a codon.

A third similar case is that of stop-suppressor mutations. Stop-suppressor mutations involve tRNAs being mutated to match stop codons. However, these tRNAs are in competition with the release factors. This prevents a stop suppressor from actually reassigning a codon because it will still be interpreted as a stop codon the vast majority of the time. The stop codons in these cases are not actually reassigned; they are simply occasionally translated as an amino acid. As such, these mutations do not provide a simple mutation that explains a stop codon reassignment.

However, stop-suppressor mutations would explain genomes that tolerate stop codons in genetic sequences while still also translating them as stop codons. Research shows that a number of non-ciliate protists with reported alternative codes avoid using the reassigned codons in crucial proteins or show a general bias against the reassigned codons [31–33, 35, 36], suggesting that using those codons is deleterious. Most likely, the codons are still sometimes interpreted as stop codons or are at least relatively inefficient in their new roles. Some of these protists have "reassigned" all of the stop codons to amino acids; the protists function because, despite this reassignment, they will nevertheless still interpret some of those codons as stop codons [31, 32]. All of this suggests that these organisms still interpret the standard stop codons as stop codons; they simply have very high read-through rates.

These findings for non-ciliates differ from those in

ciliates. Many ciliates also reassign stop codons. Some ciliates even reassign all stop codons but have been observed to use a carefully controlled system in which canonical stop codons are interpreted as stop codons near the end of RNA transcripts [41, 42]. Unlike the non-ciliate protists, they do not appear to have merely increased their read-through rates. Instead, they have either reassigned codons entirely or adopted a sophisticated mechanism to differentiate different meanings for the same codon.

Table 3 summarizes the difficulties faced by each observed reassignment. The reassignments that do not face any of the three issues identified are those that are most readily explained by mutations.

There is another class of reassignments that have not been considered: the introduction of a new stop codon. We do not know what changes would be required to recognize a new stop codon. It would appear to require changes to the release factors, but release factors are much less well understood than tRNAs. Some researchers were able to modify release factors from *Euplotes*—which normally decodes UGA as cysteine instead of a stop codon—to recognize UGA as a stop codon [75]. However, this required two mutations to the crucial domain for recognizing stop codons, replacing another domain with the version from humans and introducing an additional release factor from humans [75]. This suggests that modifying the release factor to recognize a new stop codon would be difficult, even to enable one of the canonical stop codons. Recognizing an entirely new stop codon is probably more difficult.

It should be emphasized that our purpose is not to argue that it is impossible that reassignments other than those involving a single, simple mutation could happen evolutionarily. Certainly, more complex scenarios could be envisioned that might enable a more complex reassignment. Rather, the reassignments considered in this section are the easiest reassignments to explain. Within our framework, these are the reassignments most likely to be evolved rather than designed.

3.4 Endosymbionts

While not fatal, the mutations discussed in the previous section would still be deleterious. A change could avoid being deleterious if the relevant codon was not used. However, the complete disappearance of a codon is unlikely outside of rare codons in mitochondrial genomes. Instead, any plausible account of code evolution requires the fixation of at least one deleterious mutation.

Evolutionary scenarios that require fixing a strongly deleterious mutation are generally implausible. However, there is a special case: endosymbionts, organisms that live in or on other organisms, are subject to ineffective natural selection. Nicholson et al. explain this [76]:

> Although the mechanism of genome decay is not entirely clear, it appears to primarily stem from frequent genetic drifts. Because parasites live in small, asexual and genetically bottle-necked populations, they cannot effectively eliminate deleterious mutations that sporadically occur during DNA replication. This causes irreversible accumulation of deleterious mutations and reduction of parasite genomes. Thus, it is not that parasites lose only those genes that are no longer essential for their survival in the intracellular context. It is that parasites populations cannot effectively eliminate sporadic deleterious mutations, causing accumulation of these mutations throughout their genomes, including their most essential genes.

These endosymbionts accumulate mutations in their most essential genes, including their tRNAs and release factors, causing them to change how they interpret their code. We expect, therefore, that codes will only evolve in these endosymbiont genomes.

Several variant codes are found in bacterial endosymbionts. *Stamera capleta* [27], *Hodgkinia cicadicola* [28], *Nasuia deltocephalinicola* [29] and *Zinderia quadrilineatus* [30] are all symbionts living in the digestive tract of insects. *Anaerococcus* is commonly found as part of the human microbiome and is associated with several infections. *Peptacetobacter* is found in the digestive tract of various animals. There are two orders (Mycoplasmatales and Entomoplasmatales) of Mollicutes bacteria that are parasites of various animals and plants and follow a variant code.

Two groups of Bacillus bacteria were inferred to utilize variant codes by a computational screen analyzing an extensive collection of bacterial genomes [26]. One unclassified bacterial group is explicitly stated to be from "fecal metagenomes of baboons or humans [26]". The accession IDs from the supplementary data suggest a similar origin is true for the other groups [77–80]. This suggests that these groups are likewise endosymbionts.

Members of Absconditabacteria are sometimes described as parasitic, but they are better described as predatory. (They eat and consume other bacteria [81, 82].) Gracilibacteria are also sometimes described as parasitic due to having a reduced metabolism, but they are poorly understood. Research has shown that they have a limited metabolism and are associated with *Colwellia*, a genus of bacteria [83]. They appear to live off of *Colwellia* in some way, although the details are unclear. Both groups appear not to be symbionts of plants or animals but instead live off of other bacteria.

Several variant codes are found in protist endosymbionts. *Amoeboaphelidium protococcarum* is an algal parasite [84]. *Iotanema spirale* is an endobiotic protist isolated from gecko feces [85]. The subfamily Hexamiti-

Table 3: Summary of mutational simplicity for bacterial, eukaryotic nuclear and eukaryotic mitochondrial reassignments.

Reassignment	Genome	Near-cognate tRNAs	Redundant tRNAs	Limited competition	Simple Mutation
UGA (Stop → Trp)	Bacterial	Yes	Yes	Yes	Yes
UGA (Stop → Gly)	Bacterial	Yes	Yes	No*	No*
CGG (Arg → Trp)	Bacterial	Yes	Yes	Yes	Yes
CGG (Arg → Gln)	Bacterial	Yes	Yes	Yes	Yes
CGR (Arg → Trp)	Bacterial	No	Yes	No	No
AGG (Arg → Met)	Bacterial	Yes	Yes	Yes	Yes
UAA (Stop → Lys)	Nuclear	Yes	Yes	No*	No*
UAG (Stop → Glu)	Nuclear	Yes	Yes	No*	No*
UAG (Stop → Gln)	Nuclear	Yes	Yes	No*	No*
UAG (Stop → Leu)	Nuclear	Yes	Yes	No*	No*
UAR (Stop → Glu)	Nuclear	Yes	Yes	No*	No*
UAR (Stop → Tyr)	Nuclear	Yes	Yes	No*	No*
UAR (Stop → Gln)	Nuclear	Yes	Yes	No*	No*
UGA (Stop → Trp)	Nuclear	Yes	Yes	Yes	Yes
UGA (Stop → Cys)	Nuclear	Yes	Yes	No*	No*
CUG (Leu → Ala)	Nuclear	No	Yes	Yes	No
CUG (Leu → Ser)	Nuclear	No	Yes	Yes	No
UAA (Stop → Tyr)	Mitochondrial	Yes	Yes	Yes	Yes
UAG (Stop → Tyr)	Mitochondrial	Yes	Yes	Yes	Yes
UAG (Stop → Leu)	Mitochondrial	Yes	No	No	No
UAG (Stop → Ala)	Mitochondrial	No	No	No	No
UAR (Stop → Tyr)	Mitochondrial	Yes	Yes	No	No
UGA (Stop → Trp)	Mitochondrial	Yes	Yes	Yes	Yes
CUN (Leu → Thr)	Mitochondrial	No	No	No	No
CUN (Leu → Ala)	Mitochondrial	No	No	No	No
CGG (Arg → Leu)	Mitochondrial	Yes	No	Yes	No
CGN (Arg → Gly)	Mitochondrial	Yes	No	No	No
AUA (Ile → Met)	Mitochondrial	Yes	Yes	No	No
AAA (Lys → Asn)	Mitochondrial	Yes	Yes	No	No
AGG (Arg → Lys)	Mitochondrial	Yes	No	Yes	No
AGR (Arg → Gly)	Mitochondrial	Yes	No	No	No
AGR (Arg → Ser)	Mitochondrial	Yes	Yes	No	No
AGR (Arg → Met)	Mitochondrial	No	No	No	No

* These mutations work to produce stop suppression not codon reassignment.

nae is mixed, containing both parasitic and free-living protists. However, the most parsimonious account requires that the free-living members of this taxon reverted to a free-living lifestyle from parasitic ancestors [86]. Streblomastix is a symbiont found in the gut of wood-eating termites [87]. *Rhizaria* sp. exLh and *Amoebophrya* sp. ex *Karlodinium veneficum* are parasites named after their host species.

The ciliates that exhibit various stop codon reassignments are not generally endosymbionts. Some orders of green algae also reassign some stop codons. The algae genus *Blastophysa* listed in Table 2 is endophytic [88]. However, it shares its alternative code with the related order Cladophorales, and taken together the group is not characteristically endosymbiotic.

Some families of yeast reassign the CUG codon. While there are some symbionts included in some of these groups, the groups are not characteristically symbionts.

Mitochondria, by their nature, are similar to symbionts. They live inside the cells of eukaryotes and are subject to many of the same effects as symbionts. A tendency to accumulate mutations has also been observed in mitochondria [89–91]. As such, we expect it to be possible to evolve a variant code in mitochondria even if the larger organism were not a symbiont. Nevertheless, mitochondrial variant codes might be more likely in parasitic organisms. Indeed, parasitic nematodes in the genus *Radopholus* have been observed to follow a variant code in their mitochondria [59].

Our purpose here is not to argue that it is impossible for a genetic code to evolve in a non-endosymbiotic organism. One can propose more complex scenarios that would enable another kind of organism to undergo a codon rearrangement, but endosymbionts present the easiest case to explain evolutionarily. Within our framework, these are the reassignments most likely to be evolved rather than designed.

3.5 Low Codon Usage

If codes evolve in symbionts, this explains why they can evolve despite being deleterious. Nevertheless, we expect that the changes cannot be too deleterious, or even the relaxed natural selection found in symbionts would eliminate them. The degree of deleteriousness of a code-changing mutation is primarily determined by the number of codon positions affected. A mutation affecting the translation of many codon positions is likely to be highly deleterious and even fatal. However, if relatively few codon positions are affected, the deleteriousness is minimized. Consequently, we expect evolved codes to involve the reassignment of codons that have relatively few uses. However, designed variant codes might have reassigned codons that would otherwise have many uses.

We can identify two classes of reassignments that seem unlikely to involve codons with sufficiently low usage. The first is reassignments involving multiple codons. Many reassignments involve changing the meaning of multiple, related codons. The rarity of similar codons is often correlated, but the more codons are reassigned, the greater the disruption of reassigning those codons. The easiest case, evolutionarily, will be reassigning a single codon. The second is any reassignment involved in a eukaryotic genome. Eukaryotes have an order of magnitude more genes than bacteria [92, 93]. We know that bacteria like *E. coli* can survive with a stop codon disabled [73]. However, this is much less plausible for eukaryotes.

However, there is an exception to these two classes: those non-ciliate protists that we argued previously engaged in increased read-through rather than a codon reassignment. Since the meaning of codons is not actually reassigned in these genomes, the logic above does not apply. Instead, the organisms may survive the increased read-through as long as the rate of read-through is sufficiently low.

Another relevant issue is the frequency of a particular codon. Some codons are used much less frequently than others. Obviously, it would be easier to reassign a rarely used codon. However, it is difficult to distill this into a simple criterion because the rarity of codons differs strongly from genome to genome, and there is no clear way to define what should count as rare. Consequently, we will have to leave the relative rarity of a codon out of our consideration.

Our purpose is not to argue that it is impossible to reassign a codon with higher usage. Certainly, more complex scenarios can be postulated to account for how a codon could be reassigned despite being heavily used. Instead, our point is that these are reassignments are most likely explained as evolved rather than designed in the context of our framework.

3.6 Simple Distributions

In an evolutionary code change, we would expect the distribution of a variant code to follow the phylogeny. We would expect that an organism underwent a code change at some point, and now all its descendants utilize that alternative code. Crucial to this conclusion is the fact that any sort of successful genetic code modification is extremely rare. Most mutations that modify the interpretation of the genetic code will make the organism much less fit, and they will be eliminated by natural selection. Consequently, invoking multiple genetic code modifications in closely related taxa is highly implausible. Given the large number of organisms in life's history, we are likely to see other genetic code changes, but we expect to find them in distantly related organisms.

For designed code changes, we do not expect the distribution of the code to follow the phylogeny precisely.

Some degree of fit to an evolutionary phylogeny is expected under the dependency graph model. However, we also expect some deviation from that pattern. Similar organisms would likely benefit from similar changes to the standard code. However, that would not be expected to fit a phylogenetic pattern exactly. The pattern would instead be more complex. See previous work on the dependency graph model for more exploration of the patterns expected under that model [13].

Absconditabacteria and Gracilibacteria both follow UGA (Stop → Gly), but Absconditabacteria also follows CGR (Arg → Gln), which requires two code changes early in the evolution of these phyla. The work that inferred that Absconditabacteria follows CGR (Arg → Gln) also found some evidence that CGG had also been reassigned in closely related Gracilibacteria species. They wrote:

> This may reflect a complicated history of CGG reassignment and possible reversion to arginine translation. [26]

The distribution of variant codes in ciliates is complex (see Figure 1). The so-called ciliate code, which follows UAR (Stop → Gln), originates four different times. A "stopless" code, which reassigns all three stop codons, originates three times. Three other codes that reassign stop codons are also found in the ciliates: UAR (Stop → Glu), UAR (Stop → Tyr) and UGA (Stop → Cys). In addition, an unclassified Oligohymenophorea species has been observed to follow UAA (Stop → Lys) and UAG (Stop → Glu). The distribution of these codes in ciliates is complex.

Another complex distribution is in a class of green algae (see Figure 2). Only one code is involved: UAR (Stop → Gln). However, it is found in three distinct groups. The paper publishing the results for this class called it a complex distribution and sought to explain it by suggesting that the code originated once but reverted in one of the orders [44].

Some yeast use a variant nuclear code with a complex distribution (see Figure 3). Three different groups reassign the CUG codon to two distinct amino acids: serine and alanine.

The distribution of mitochondrial codes within *Saccharomycetaceae* is also complex (see Figure 4). The reassignment AUA (Ile → Met) shows up in two distinct groups. The entire family reassigns the CUN block to two different amino acids.

The distribution of mitochondrial codes within the order Sphaeropleales is also complex (see Figure 5). Three distinct groups follow AGR (Arg → Ala). One family follows AGR (Arg → Met). Two groups reassign the UAG stop codon, one to alanine and the other to leucine. Two groups follow CGG (Arg → Leu).

The distribution of mitochondrial codes within metazoans is especially complex (see Figure 6). Most bilaterian metazoans follow AUA (Ile → Met), but the exceptions cut through the taxonomy. A simplistic evolutionary interpretation would suggest that AUA (Ile → Met) occurred in no fewer than *five* distinct places in the metazoan phylogeny. The transition ended up occurring in all but a couple of lineages. This seems excessively unparsimonious, suggesting that we might instead propose, still with an evolutionary framework, that the AUA (Ile → Met) transition occurred at the root of the bilaterian animals. The change was then reverted twice, once in the Platyhelminthes and once in the common ancestor of Echinodermata and Hemichordata. This account may be more parsimonious, but with a single change and two reversions, it is not much better.

Furthermore, of the three bilaterian metazoan phyla that do not follow AUA (Ile → Met), two follow AAA (Lys → Asn). These two reassignments come in the two distinct lineages that must have reverted the AUA (Ile → Met) reassignment. As such, in an evolutionary interpretation, both lineages must have experienced a new reassignment shortly after reverting the previous reassignment.

Most bilaterian metazoans follow AGR (Arg → Ser). Additionally, two classes of sponges independently follow this reassignment. In the chordates, one lineage, Cephalocordata, maintains that assignment; another, Tunicata, follows AGR (Arg → Gly), and the last, Vertebrata, follows AGR (Arg → Stop).

The most complex distribution belongs to the arthropod mitochondrial genomes (see Figure 7). Many follow AGG (Ser → Lys), but the distribution is chaotic. An evolutionary account of the distribution requires at least 20 changes, with the code repeatedly reverting back and forth between decoding AGG as serine or lysine.

Other genetic code changes are only distantly related to each other. The four individually listed bacterial species in Table 1 all belong to the phylum Pseudomonadata but belong to three classes and four different orders. Likewise, *Anaerococcus* and *Peptaceobacter* both belong to the order Bacillota but to different classes. The variant codes found in non-ciliate protists are mostly found in different phyla with one exception—the phylum Fornicata—which contains two variant codes but that are found in different classes. In all of these cases, the variant codes are found in distantly related organisms, which makes sense if they are the outcome of a rare modification to the genetic code in an evolutionary scenario.

It should be intuitively clear that the given examples defy evolutionary expectations with regards to the distribution of the genetic code. However, for the purposes of defining a criterion, we use a more precise test. We consider taxa to be closely related if they have the same taxonomic rank and a common major parent taxon. Thus, all orders that belong to a particular class are considered to be closely related. We consider the dis-

Figure 1: The phylogeny of ciliates with nuclear genetic codes. **doi:** 10.5048/BIO-C.2024.1.f1

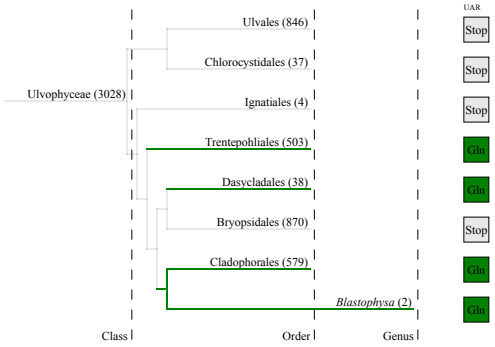

Figure 2: The phylogeny of Ulvophyceae, a class of green algae with nuclear genetic codes. **doi:** 10.5048/BIO-C.2024.1.f2

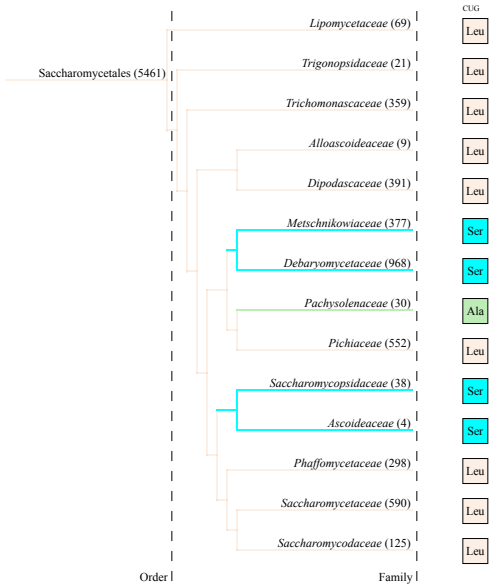

Figure 3: The phylogeny of Saccharomycetales with nuclear genetic codes. **doi:** 10.5048/BIO-C.2024.1.f3

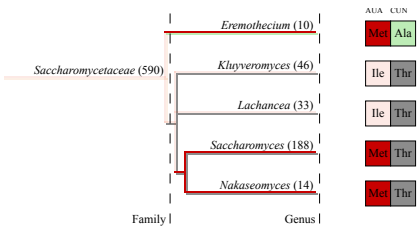

Figure 4: The phylogeny of Saccharomycetaceae with mitochondrial genetic codes. **doi:** 10.5048/BIO-C.2024.1.f4

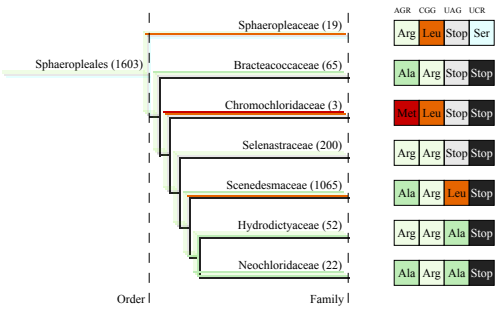

Figure 5: The phylogeny of Sphaeropleales with Mitochondrial genetic codes. **doi:** 10.5048/BIO-C.2024.1.f5

tribution to be complex if multiple genetic code modifications are necessary to account for the distribution of codes amongst these closely related taxa. Given the rarity of genetic code modifications, it should be highly unlikely that any two such closely related taxa would both happen to undergo one.

Almost all of the examples given above pass this test. The exception is a few cases in the ciliates' tree, namely the order Cyclotrichida and the families Cyrtolophosididae and Nassulidae. While arguably part of the overall pattern of complexity in the distribution of variant genetic codes within ciliates, they are not closely related to other taxa with variant codes, being the only taxon with a known variant code within their class.

Our purpose here is not to argue that evolutionary theory must follow consistent phylogenetic patterns. More complex scenarios can be postulated that would account for a deviation from a simple, tree-like pattern. Indeed, the loss-driven codon reassignment model attempts to explain the propensity for multiple genetic code mutations among closely related taxa [12]. The evaluation of the plausibility of such scenarios is beyond our scope. The point here is that, within our framework, we would expect an evolutionarily derived code to follow a simple distribution and many designed codes to follow a complex distribution.

3.7 Correlation Between Criteria

The previous sections describe five different criteria by which we expect designed codes to differ from evolved codes. We have seen that some reassignments are characteristic of high-level taxa, while others are characteristic of low-level taxa. Some reassignments would require relatively simple modifications to the translation machinery, while others would require complicated changes. Some reassignments are found in endosymbiotic organisms, while others are found in free-living organisms. Some reassignments involve rare codons, while others involve common codons. Some reassignments follow phylogenetic expectations, while others deviate from them.

If our framework is correct, we would expect that

Table 4: Fisher's one-tailed exact test p-values for correlation between different criteria. Each reassignment listed in Table 2 is counted as a data point.

	Low Level Taxon	Simple Mutation	Endosymbiont	Rare Codons	Simple Distribution
Low Level Taxa		3.65×10^{-5}	1.50×10^{-7}	1.10×10^{-3}	6.37×10^{-4}
Simple Mutation	3.65×10^{-5}		1.17×10^{-8}	7.28×10^{-7}	1.32×10^{-10}
Endosymbiont	1.50×10^{-7}	1.17×10^{-8}		1.68×10^{-10}	1.17×10^{-8}
Rare Codons	1.10×10^{-3}	7.28×10^{-7}	1.68×10^{-10}		1.24×10^{-5}
Simple Distribution	6.37×10^{-4}	1.32×10^{-10}	1.17×10^{-8}	1.24×10^{-5}	

Table 5: The different criteria compared to a proposed classification. Each reassignment listed in Table 2 is counted as a data point.

	Low Level Taxon	Simple Mutation	Endosymbiont	Rare Codons	Simple Distribution
Fisher's exact test p-value	1.77×10^{-11}	1.14×10^{-13}	2.03×10^{-10}	9.71×10^{-8}	2.52×10^{-10}
Agreement	93.5%	85.6%	97.5%	67.8%	76.7%
Designed codes	8.7%	21.0%	4.3%	46.8%	33.9%
Evolved codes	100.0%	100.0%	100.0%	100.0%	100.0%

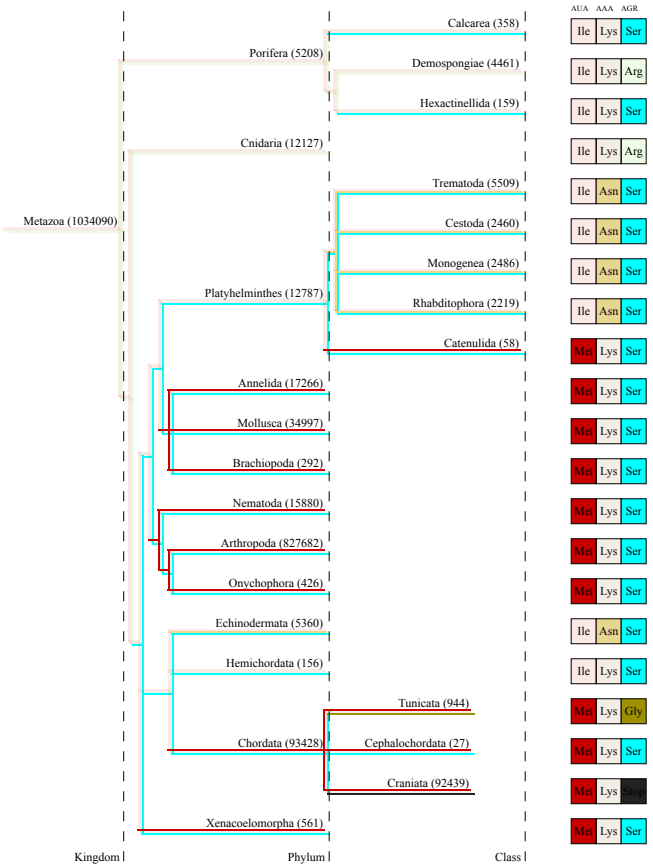

Figure 6: The phylogeny of Metazoa with Mitochondrial genetic codes. **doi:** 10.5048/BIO-C.2024.1.f6

these different criteria would agree on whether a codon reassignment is designed or evolved. We do not expect exact agreement because our criteria are all somewhat approximate. Table 4 shows the p-values from Fisher's exact test for the correlation between each of the criteria. We find that all of the criteria correlate with each other with a highly statistically significant level of support.

Each reassignment listed in Table 2 is counted as a single data point. Multiple, unrelated codon reassignments within a single taxon are counted as multiple, distinct reassignments. However, when a taxon reassigns related codons or when multiple taxa which are considered to share a common ancestor all follow a variant code, these are counted as a single reassignment. The chaotic distribution of the arthropod codes is counted as a single entry.

It is also useful to compare all of the criteria with a proposed classification of observed codes as either evolved or designed. For that proposed classification, we classify codes as evolved if and only if they are deemed evolved by all five criteria. That is, codes that are found in low-level taxa, require only a simple mutation, are found in symbionts, reassign rare codons, and have a simple distribution are deemed to have evolved. All other codes are deemed designed. We take this approach because the criteria are defined such that every evolved code should follow them, but for most criteria, it would not be surprising if some designed codes also followed them. For example, an evolved code would be expected to only reassign a rare codon, but a designed code may or may not reassign a rare codon.

Table 5 presents the comparison of each criteria to this standard. All criteria show a strongly statistically significant correlation with that classification. The criteria agree with the overall classification from 67.8% to

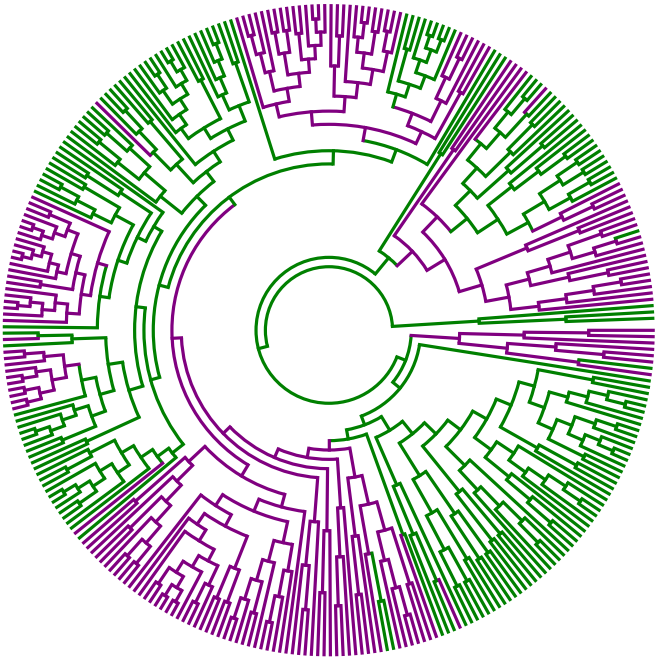

Figure 7: A phylogeny of arthropod species. Green represents lineages that decode AGG as serine in the mitochondria. Purple represents lineages that decode AGG as lysine in the mitochondria. The inferred code for individual species is taken from Abascal et al [57]. The code for ancestral lineages and taxa without an inferred code is inferred by parsimony. **doi:** 10.5048/BIO-C.2024.1.f7

97.5% of the time. Due to the way we defined the classification, evolved codes exhibit each of the properties 100% of the time. However, there is considerable variation with respect to how many designed codes fit each criterion. Partially, this is because, as mentioned, many of the criteria are such that a designed code may or may not fit them. We will look more closely at some of the misfitting examples in Section 3.8.

3.8 Misfits

3.8.1 Introduction

While we show good overall correlation among the criteria and between the criteria (Table 4) and an overall classification (Table 5), there are some variant codes that fit less well. A correct theory must be correct in every case, and, thus, simply pointing out the overall fit is insufficient. We have to account for the cases that do not fit as well into our framework. At the same time, our criteria are approximations. It is thus to be expected that some cases will not fit as well. We will find that the misfits are fairly explicable.

3.8.2 Saccharomycetaceae

One of these misfits is the yeast family *Saccharomycetaceae*, which contains three different mitochondrial code reassignments. These reassignments look designed according to most of our criteria: they require a complex mutation, involve common codons and have a complex distribution. However, they are found between the levels of family and genus—low-level taxa. These reassignments are responsible for all mismatches between the low-level taxonomic criteria and the proposed classification.

These results can be reconciled with the framework if we conclude that *Saccharomycetaceae* does not share a single common ancestor. Recall that we proposed the idea that limited common ancestry applied below the level of family as an approximation. Indeed, prior thought has suggested that, in some cases, genera rather than families are the independently designed organisms [62]. This would imply that the limited common ancestry in this family began with at least three species corresponding to the three distinct mitochondrial codes.

3.8.3 UGA (Stop → Trp)

The most common reassignment is UGA (Stop → Trp). Most criteria suggest that it could have evolved: it can be produced by a simple mutation, involves a rare codon and has a simple distribution. It is found either in endosymbiotic bacteria or mitochondria—but in both very high-level and very low-level taxa. In our proposed classification, these codes are classified as evolved or designed simply on the basis of their taxonomic rank.

We propose that this is a consequence of this reassignment being both evolutionarily achievable and often useful. This reassignment simplifies the code by eliminating one of the two exceptions where a codon ending with guanine has a different meaning than the corresponding codon ending with adenine. For bacteria and mitochondria, one release factor recognizes UAA and UAG while another recognizes UAA and UGA. Eliminating UGA as stop codon means that only one of those release factors is necessary.

There is a pattern to the distribution of UGA (Stop → Trp) within multicellular life. Photosynthesizing multicellular organisms, algae and plants do not, in general, have mitochondria that follow UGA (Stop → Trp). Non-photosynthesizing multicellular organisms, animals, and fungi almost all have mitochondria that follow UGA (Stop → Trp). The photosynthesizing organisms rely on chloroplasts for much of their ATP production and do not have the same requirements for their mitochondria. As such, we find this change in the mitochondria in complex non-photosynthesizing organisms that would have the most need for efficiency.

A particularly notable case is that of the bacterial orders Mycoplasmatales and Entomoplasmatales. These are the only cases where a designed variant code is attributed to a non-mitochondrial endosymbiont genome. Without it, we would have 100% agreement between the endosymbiont criteria and our proposed classification. Orders are considered high-level taxa. However, these are relatively small orders, and one could propose that this order is related by limited common descent. If so,

Figure 8: A comparison of the standard codon table and UGA (Stop → Trp). **doi:** 10.5048/BIO-C.2024.1.f8

Figure 9: A comparison of the standard codon table and AUA (Ile → Met). **doi:** 10.5048/BIO-C.2024.1.f9

the low-level taxa criterion would be in error, and this would actually be an evolved rather than a designed code.

3.9 Design Tradeoffs

If many of these variant codes are designed, this will imply some design reason for the variation. The reason may not always be apparent. Nevertheless, we have some indications of possible reasons for some of the code variants. At the same time, we must acknowledge that the ideas presented here are more speculative.

Ciliates have an unusual genomic architecture. In particular, ciliates have two nuclei: a micronucleus and a macronucleus. The micronucleus carries the genetic information from generation to generation, and the macronucleus is generated from the micronucleus by amplification and heavy editing. Given this unusual architecture, it is plausible that the trade-offs for the genetic code design would be different than in a typical cell, thus explaining the prevalence of non-canonical codes among the ciliates.

Mitochondria have very small genomes and sometimes exist in large numbers inside eukaryotic cells. As such, it makes sense that the trade-offs differ for mitochondria compared to nuclear or bacterial genomes. In particular, it would likely make sense to simplify the codon table. Two common mitochondrial changes, UGA (Stop → Trp) (see Figure 8) and AUA (Ile → Met) (see Figure 9), simplify the standard genetic code by creating split family boxes. We previously considered how this reassignment is associated with complex non-photosynthesizing life in Section 3.8.3. Likewise, another common change, AGR (Arg → Ser), simplifies the code by creating a family box (Figure 11).

AUA (Ile → Met) (see Figure 9) and AAA (Lys → Asn) (see Figure 10) make opposite changes to the code. AUA (Ile → Met) simplifies a box into two split family boxes. AAA (Lys → Asn) takes two split family boxes and makes a three-one box. This is interesting because AAA (Lys → Asn) is found in those metazoan mitochondrial lineages that do not follow AUA (Ile → Met). This would be explained if combining opposite changes in these variants would not make sense.

3.10 Conclusions

We have presented a framework for understanding the character and distribution of variant genetic codes. That framework implies the existence of two classes of codes: designed and evolved. We have found that the known variant codes fit well into these two categories. Furthermore, the framework provides excellent explanatory power for the properties of these two classes.

Evolved codes are only found in low-level taxa. Evo-

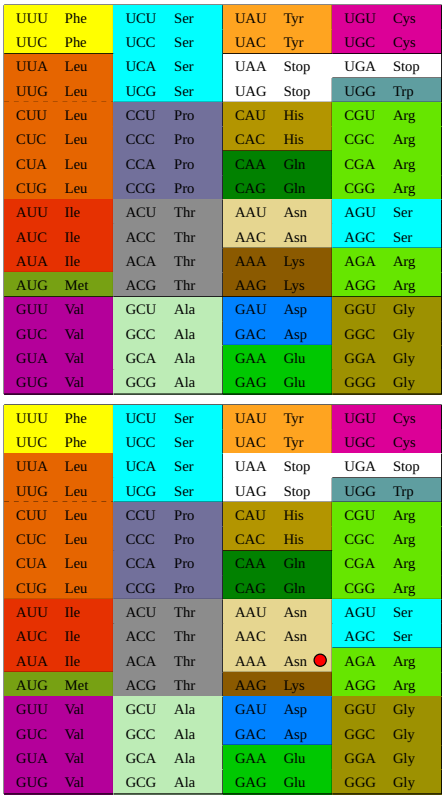

Figure 10: A comparison of the standard codon table and AAA (Asn → Met). **doi:** 10.5048/BIO-C.2024.1.f10

Figure 11: A comparison of the standard codon table and AGR (Arg → Ser). **doi:** 10.5048/BIO-C.2024.1.f11

lutionary changes can be easily explained by simple mutations. They are only found in endosymbiont genomes, which have properties that can explain the fixation of deleterious mutations. The codons that are reassigned are consistently rare. The distribution of such changes is simple and follows phylogenetic expectations.

On the other hand, designed codes are found in high-level taxa of at least genus-level but typically higher. They involve many reassignments that are difficult to explain with any sort of simple mutation. They are found in free-living organisms. They sometimes reassign codons that are expected to be rare. They are often distributed in a complex fashion that does not fit phylogenetic expectations.

In some cases, we have some idea of why a designer would make particular adjustments. Ciliates may be using alternative codes to adapt to their unusual use of multiple nuclei. Mitochondria, especially those in non-photosynthesizing multicellular organisms, appear to have had their code streamlined. Considerable variation is found in metazoans, which have the most to gain from optimized mitochondria.

However, new variant codes are continuing to be discovered, and known codes are continuing to be fully explored. Indeed, new ciliate codes were reported while this paper was under review [38]. The real test of our framework will be in future discoveries. Some possible reassignments would be very difficult for the framework to accommodate, such as an evolutionarily difficult reassignment in a very low-level taxon. Most dramatically, the observation of the evolution of an evolutionary difficult code would probably be fatal for this framework. Others code reassignments could be classified inconsistently by different criteria, and so undermine the correlation among the different criteria proposed in our framework. While the identified misfits seem explicable, additional misfits could be identified that defy explanation. We will have to wait for that future research to see how our framework fares.

4. EVOLUTIONARY THEORY

4.1 Introduction

As discussed in the introduction, evolutionary theory might appear, on first inspection, to have good explanatory power for the character and distribution of variant genetic codes. There are three different aspects to this apparent explanatory power. Firstly, it would explain the near universality of the genetic code. Whereas a designer might have chosen to use very different codes in different organisms, it would be difficult for evolution to account for such varied codes. As such, a near-universal

genetic code is a prediction of common descent. Secondly, it would explain the prevalence of variant codes in mitochondria and other simple genomes. While a designer might choose to use any code in any organism regardless of the complexity of the genome, evolution appears constrained to modifying only the codes of simple genomes. As such, it is a prediction of common descent that variant codes would be found in these simple genomes. Thirdly, it would explain why variant codes are distributed according to phylogenetic patterns. While a designer might choose to use variant codes in any sort of pattern, evolution is restricted to following phylogenetic distributions. Thus, common descent predicts that variant codes follow a phylogenetic distribution.

The repeated idea is that evolution is constrained: there are only certain kinds of code changes and patterns that can appear. As such, evolutionary theory predicts those patterns. Insofar as those patterns are found in nature, the predictions of evolutionary theory have been confirmed. However, do these predictions follow from evolutionary theory, and are the predictions successful?

4.2 Universality

The universality of the genetic code is sometimes claimed as an important and successful prediction of evolutionary theory [6]:

> More recently, molecular genetics has demonstrated a particularly dramatic unity—the genetic code. If organisms had arisen independently they could perfectly well have used different codes to connect the 64 trinucleotide codons to the 20 amino acids; but if they arose by common descent any alteration in the code would be lethal, because it would change too many proteins at once. Hence the finding of the same genetic code in microbes, plants, and animals (except for minor variations in intracellular organelles) spectacularly confirms a strong evolutionary prediction.

Today, the claimed prediction must be modified to account for the much more extensive known variation in the genetic code. The evolutionary theorist will argue that the code is close enough to being universal that it is still a pretty good confirmation of evolutionary theory.

It is not always appreciated, but the evolutionary prediction depends on the assumption that the universal genetic code was well established in the last universal common ancestor. This assumption is not actually expected under evolutionary theory. Under an evolutionary account, the genetic code did not arise all at once but gradually developed over time. It is inescapable that there would have been different organisms with different codes. It is unexpected that none of the diversity of life that we currently know is derived from any of those organisms with those alternative codes.

The argument that the code should not vary because alterations to the code would be highly deleterious was published in a couple of 1963 papers, both of which made this point. The first paper states it as follows [94]:

> All this supposes that at some very primitive stage all organisms had the same code. This might have happened if the system arose only once, though it is not obvious that the code should have evolved by a whole series of distinct additions to a simpler system without diversification occurring.

The second paper agrees [95]:

> If the code is not universal, the number of different codes should represent the number of different primordial ancestors that either existed during the time the present code was being completed, or existed when organisms were so simple that changes in practically all proteins were not always fatal. In either case, if different codes do exist they should be associated with major taxonomic groups such as phyla or kingdoms that have their roots far in the past.

In 1973, Crick and Orgel put forward the universal genetic code as evidence for directed panspermia [96]:

> Several orthodox explanations of the universality of the genetic code can be suggested, but none is generally accepted to be completely convincing. It is a little surprising that organisms with somewhat different codes do not coexist. The universality of the code follows naturally from infective theory of the origins of life on earth would represent a clone derived from a single extraterrestrial organism even if many codes were represented at the primary site where life began, only a single one might have operated in the organisms used to infect the Earth.

These authors thought it was surprising that there was a single universal genetic code but believed it could be explained if all of life descended from a single cell that already had a fully established code.

Therefore, evolutionary theory does not predict a single universal genetic code. If anything, evolutionary theory would favor the idea that the genetic code was still in flux at the time of the last universal common ancestor. Thus, we should see somewhat divergent codes in the highest-level taxa. The idea that the last universal common ancestor would already have a fully established genetic code is not expected under evolutionary theory.

The assumption is added to account for the observation of an early universal genetic code.

Sometimes these quotations are invoked in an attempt to show that these theorists predicted the existence of variant codes. However, they did not predict the sort of variant codes that we actually observe. Their predictions were based on the idea that there would be codes surviving from the time of the evolution of the genetic code. What we observe instead are modifications of the standard code. They are not associated with the high-level taxa as predicted by these authors.

It is not that evolutionary theory is unable to accommodate a nearly universal genetic code. It is possible that all other lineages died off, leaving only those carrying the now canonical genetic code. However, evolutionary theory could have at least as easily accommodated numerous different codes. Despite the claims of evolutionary theorists, this is not a prediction of universal common ancestry and thus does not provide strong evidence for it.

4.3 Genomes with Variant Codes

Mitochondrial variant codes are much more widespread than nuclear or bacterial variant codes. Upon first inspection, this seems to make sense within the evolutionary framework because it would be much easier to evolve a variant code in mitochondria with their tiny genome.

There are approximately three times as many mitochondrial variants as nuclear code variants (see Table 2). But this is not actually why mitochondrial variants are much more widespread. In an evolutionary framework, mitochondrial code changes took place in what would become the common ancestors of highly diverse groups, particularly animals and fungi. The widespread use of variant mitochondrial codes is due to the placement of those code changes in early lineages rather than the propensity of mitochondrial genomes to evolve.

Furthermore, there is not enough difference between the amounts of mitochondrial and nuclear code evolution for the evolutionary account to work. A mitochondrion has on the order of ten genes, but a free-living bacterium has on the order of one thousand, a hundredfold difference. We would expect that reassigning a codon would be exponentially more difficult as the number of genes increased. If the small genomes of mitochondria are responsible for their propensity to undergo code changes, the difference should be much more pronounced. There should be numerous mitochondrial variations for every nuclear variation, not simply a three-to-one ratio.

Furthermore, variant codes are found in nuclear genomes that are not particularly small. They are found in ciliates, which have comparable numbers of genes to the human genome. Additionally, we find them in some multicellular green algae. In fact, we find more code variation in eukaryotic nuclear genomes than in bacterial genomes, despite eukaryotes having much larger genomes. Despite the initial impression, evolutionary theory does not account well for the kinds of genomes with variant codes.

4.4 Phylogenetics

Evolutionary theory's clearest prediction is that the distribution of a code should follow the phylogeny. In NCBI's list of variant codes, codes are named after the taxonomic groups that exhibit them, such as ciliates, echinoderms, flatforms, vertebrates, invertebrates and fungi. This gives the impression that these codes neatly fit into a standard phylogeny. A cursory look at the metazoan mitochondrial codes gives the impression of a nested hierarchy of codes.

However, as we saw when we looked at the complex distribution of codes, this was not the case. In many cases, the distribution of a code is complex, defying evolutionary explanation. Codes recur in closely related groups in a way not explained by common descent. Evolutionary theory has to invoke inexplicable events such as reversions to the standard code.

Today we find ideas such as the loss-driven codon reassignment model [12] which predicts a more complex distribution of variant genetic codes. This avoids evolutionary theory being falsified by the complex distribution of these variant genetic codes. However, it does so at the cost of undermining the prediction of phylogenetic variant code distribution that seemed to follow from the theory of common descent.

5. CONCLUSIONS

We have put forward a framework for understanding the character and distribution of variant genetic codes. This framework is based on three tenets. Firstly, the genetic code is well-engineered and suited for most genomes. Secondly, some genomes benefit from a different genetic code. Thirdly, some genomes have accumulated mutational errors and thus interpret their genetic code differently.

We surveyed variant genetic codes to evaluate this framework, identifying the distribution and character of all known variant codes. We found that we could distinguish the designed and evolved codes by considering organismal lifestyle, taxonomic ranks, evolutionary feasibility, multiple codon reassignments and complexity of distribution. We showed that these different criteria correlate well with each other and that deviations from the general pattern are explicable. As such, our framework provides excellent explanatory power for the known variant codes.

Initially, evolutionary theory appeared to have some explanatory power. However, upon closer inspection, the features of the variant codes that seemed well explained by evolutionary theory turned out to either be inaccurate or to not follow from evolutionary theory.

We have not attempted to critique evolutionary theory based on the mechanical difficulties involved in genetic code evolution. Certainly, models of genetic code evolution face difficulties [11, 97–99]. Some of the data presented here in terms of the difficulty of certain codon reassignments contributes to that difficulty. However, that is not our concern here. Rather, we have sought to show that even based solely on the distribution and character of the variant genetic codes, our framework is a better explanation than that of evolutionary theory. Future research will put our framework to the test. We expect more variant codes to be discovered in the future, and our theory will ultimately be tested by its ability to accommodate those discoveries.

This paper makes some important contributions to the development of common design as an alternative model to common descent in expanding the analysis from genes to the genetic code itself. It takes an area previously argued to demonstrate universal common ancestry and demonstrates a framework with superior explanatory power based on common design. This paper moves beyond simply critiquing universal common ancestry and provides its own understanding of the design and history of life. It also explicitly incorporates limited common ancestry as a core part of the framework. Altogether, this paper constitutes an important step towards the development of common design as a full-fledged alternative to universal common ancestry.

1. Steel M, Penny D (2010) Origins of life: Common ancestry put to the test. Nature 465(7295):168–169. doi:10.1038/465168a
2. Koonin EV, Wolf YI (2010) The common ancestry of life. Biology Direct 5(1):64. doi:10.1186/1745-6150-5-64
3. Theobald DL (2011) On universal common ancestry, sequence similarity, and phylogenetic structure: The sins of P-values and the virtues of Bayesian evidence. Biology Direct 6(1):60. doi:10.1186/1745-6150-6-60
4. Delaye L, Becerra A (2012) Cenancestor, the Last Universal Common Ancestor. Evolution: Education and Outreach 5(3):382–388. doi:10.1007/s12052-012-0444-8
5. Ridley M (1987) The problems of evolution. British Journal for the Philosophy of Science 38(3):412–414
6. Davis BD (1985) Molecular genetics and the foundations of Evolution. Perspectives in Biology and Medicine 28(2):251–268. doi:10.1353/pbm.1985.0051
7. Dawkins R (2010) The Greatest Show on Earth: The Evidence for Evolution. Free Press (New York)
8. Nelson P (2021) Can universal common Descent be tested? In: The Comprehensive Guide to Science and Faith: Exploring the Ultimate Questions about Life and the Cosmos. Harvest House Publishers (Eugene)
9. Meyer SC (2007) Explore Evolution: The Arguments for and against Neo-Darwinism. Hill House Publishers (Malvern, Vic, Australia)
10. Osawa S, Jukes TH (1989) Codon reassignment (codon capture) in evolution. Journal of Molecular Evolution 28:271–278. doi:10.1007/BF02103422
11. Schultz DW, Yarus M (1994) Transfer RNA Mutation and the Malleability of the Genetic Code. Journal of Molecular Biology 235(5):1377–1380. doi:10.1006/jmbi.1994.1094
12. Mühlhausen S, Findeisen P, Plessmann U, Urlaub H, Kollmar M (2016) A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. Genome Research 26(7):945–955. doi:10.1101/gr.200931.115
13. Ewert W (2018) The Dependency Graph of Life. BIO-Complexity 2018(3):1–27. doi:10.5048/BIO-C.2018.3
14. Ewert W (2023) AminoGraph Analysis of the Auditory Protein Prestin From Bats and Whales Reveals a Dependency-Graph Signal That Is Missed by the Standard Convergence Model. BIO-Complexity 2023(1). doi:10.5048/BIO-C.2023.1
15. Macosko JC, Smelser AM (2013) An Ode to the Code: Evidence for Fine-Tuning in the Standard Codon Table. In: Biological Information, Figure 1, 418–434. WORLD SCIENTIFIC. doi:10.1142/9789814508728_0018
16. Wichmann S, Ardern Z (2019) Optimality in the standard genetic code is robust with respect to comparison code sets. Biosystems 185:104023. doi:10.1016/j.biosystems.2019.104023
17. Freeland SJ, Hurst LD (1998) The Genetic Code Is One in a Million. Journal of Molecular Evolution 47(3):238–248. doi:10.1007/PL00006381
18. Freeland SJ, Wu T, Keulmann N (2003) The case for an error minimizing standard genetic code. Origins of Life and Evolution of the Biosphere 33(4/5):457–477. doi:10.1023/A:1025771327614
19. Johnson LJ, Cotton JA, Lichtenstein CP, Elgar GS, Nichols RA, Polly PD, Le Comber SC (2011) Stops making sense: Translational trade-offs and stop codon reassignment. BMC Evolutionary Biology 11(1). doi:10.1186/1471-2148-11-227
20. Itzkovitz S, Alon U (2007) The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Research 17(4):405–412. doi:10.1101/gr.5987307
21. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST (2022) Database resources of the national center for biotechnology information. Nucleic Acids Research 50(D1):D20–D26. doi:10.1093/nar/gkab1112
22. Li C, Zhang J (2019) Stop-codon read-through arises largely from molecular errors and is generally nonadaptive. PLOS Genetics 15(5):e1008141. doi:10.1371/journal.pgen.1008141
23. Biziaev N, Sokolova E, Yanvarev DV, Toropygin IY, Shuvalov A, Egorova T, Alkalaeva E (2022) Recognition of 3′ nucleotide context and stop codon readthrough are determined during mRNA translation elongation. Journal of Biological Chemistry 298(7):102133. doi:10.1016/j.jbc.2022.102133
24. Jungreis I, Lin MF, Spokony R, Chan CS, Negre N, Victorsen A, White KP, Kellis M (2011) Evidence of abundant stop codon readthrough in *Drosophila* and other metazoa. Genome Research 21(12):2096–2113. doi:10.1101/gr.119974.110
25. Del Cortona A, Leliaert F, Bogaert KA, Turmel M, Boedeker C, Janouškovec J, Lopez-Bautista JM, Verbruggen H, Vandepoele K, De Clerck O (2017) The Plastid Genome in Cladophorales Green Algae Is Encoded by Hairpin Chromosomes. Current Biology 27(24):3771–3782.e6. doi:10.1016/j.cub.2017.11.004
26. Shulgina Y, Eddy SR (2021) A computational screen for alternative genetic codes in over 250,000 genomes. eLife 10:71402. doi:10.7554/eLife.71402
27. Salem H, Bauer E, Kirsch R, Berasategui A, Cripps M, Weiss B, Koga R, Fukumori K, Vogel H, Fukatsu T,

Kaltenpoth M (2017) Drastic Genome Reduction in an Herbivore's Pectinolytic Symbiont. Cell 171(7):1520–1531.e13. doi:10.1016/j.cell.2017.10.029

28. McCutcheon JP, McDonald BR, Moran NA (2009) Origin of an Alternative Genetic Code in the Extremely Small and GC–Rich Genome of a Bacterial Symbiont. PLoS Genetics 5(7):e1000565. doi:10.1371/journal.pgen.1000565
29. Bennett GM, Moran NA (2013) Small, Smaller, Smallest: The Origins and Evolution of Ancient Dual Symbioses in a Phloem-Feeding Insect. Genome Biology and Evolution 5(9):1675–1688. doi:10.1093/gbe/evt118
30. McCutcheon JP, Moran NA (2010) Functional Convergence in Reduced Genomes of Bacterial Symbionts Spanning 200 My of Evolution. Genome Biology and Evolution 2:708–718. doi:10.1093/gbe/evq055
31. Záhonová K, Kostygov AY, Ševčíková T, Yurchenko V, Eliáš M (2016) An unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. Current Biology 26(17):2364–2369. doi:10.1016/j.cub.2016.06.064
32. Bachvaroffi TR (2019) A precedented nuclear genetic code with all three termination codons reassigned as sense codons in the syndinean Amoebophrya sp. Ex Karlodinium veneficum. PLoS ONE 14(2). doi:10.1371/journal.pone.0212912
33. Keeling PJ, Doolittle WF (1997) Widespread and ancient distribution of a noncanonical genetic code in diplomonads. Molecular Biology and Evolution 14(9):895–901. doi:10.1093/oxfordjournals.molbev.a025832
34. De Koning AP, Noble GP, Heiss AA, Wong J, Keeling PJ (2008) Environmental PCR survey to determine the distribution of a non-canonical genetic code in uncultivable oxymonads. Environmental Microbiology 10(1):65–74. doi:10.1111/j.1462-2920.2007.01430.x
35. Keeling PJ, Leander BS (2003) Characterisation of a Non-canonical genetic code in the oxymonad Streblomastix strix. Journal of Molecular Biology 326(5):1337–1349. doi:10.1016/S0022-2836(03)00057-3
36. Karpov SA, Mikhailov KV, Mirzaeva GS, Mirabdullaev IM, Mamkaeva KA, Titova NN, Aleoshin VV (2013) Obligately phagotrophic aphelids turned out to branch with the earliest-diverging fungi. Protist 164(2):195–205. doi:10.1016/j.protis.2012.08.001
37. Pánek T, Žihala D, Sokol M, Derelle R, Klimeš V, Hradilová M, Zadrobílková E, Susko E, Roger AJ, Čepička I, Eliáš M (2017) Nuclear genetic codes with a different meaning of the UAG and the UAA codon. BMC Biology 15(1). doi:10.1186/s12915-017-0353-y
38. McGowan J, Kilias ES, Alacid E, Lipscombe J, Jenkins BH, Gharbi K, Kaithakottil GG, Macaulay IC, McTaggart S, Warring SD, Richards TA, Hall N, Swarbreck D (2023) Identification of a non-canonical ciliate nuclear genetic code where UAA and UAG code for different amino acids. PLOS Genetics 19(10):e1010913. doi:10.1371/journal.pgen.1010913
39. Seah BKB, Singh A, Swart EC (2022) Karyorelict ciliates use an ambiguous genetic code with context-dependent stop/sense codons. Peer Community Journal 2:e42. doi:10.24072/pcjournal.141
40. Yan Y, Maurer-Alcalá XX, Knight R, Pond SL, Katz LA (2019) Single-cell transcriptomics reveal a correlation between genome architecture and gene family evolution in ciliates. mBio 10(6). doi:10.1128/mBio.02524-19
41. Heaphy SM, Mariotti M, Gladyshev VN, Atkins JF, Baranov PV (2016) Novel Ciliate Genetic Code Variants Including the Reassignment of All Three Stop Codons to Sense Codons in Condylostoma magnum. Molecular Biology and Evolution 33(11):2885–2889. doi:10.1093/molbev/msw166
42. Swart EC, Serra V, Petroni G, Nowacki M (2016) Genetic Codes with No Dedicated Stop Codon: Context-Dependent Translation Termination. Cell 166(3):691–702. doi:10.1016/j.cell.2016.06.020
43. Sengupta S, Yang X, Higgs PG (2007) The mechanisms of codon reassignments in mitochondrial genetic codes. Journal of Molecular Evolution 64(6):662–688. doi:10.1007/s00239-006-0284-7
44. Cocquyt E, Gile GH, Leliaert F, Verbruggen H, Keeling PJ, De Clerck O (2010) Complex phylogenetic distribution of a non-canonical genetic code in green algae. BMC Evolutionary Biology 10(1). doi:10.1186/1471-2148-10-327
45. Krassowski T, Coughlan AY, Shen XX, Zhou X, Kominek J, Opulente DA, Riley R, Grigoriev IV, Maheshwari N, Shields DC, Kurtzman CP, Hittinger CT, Rokas A, Wolfe KH (2018) Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. Nature Communications 9(1). doi:10.1038/s41467-018-04374-7
46. Groenewald M, Hittinger C, Bensch K, Opulente D, Shen XX, Li Y, Liu C, LaBella A, Zhou X, Limtong S, Jindamorakot S, Gonçalves P, Robert V, Wolfe K, Rosa C, Boekhout T, Čadež N, Péter G, Sampaio J, Lachance MA, Yurkov A, Daniel HM, Takashima M, Boundy-Mills K, Libkind D, Aoki K, Sugita T, Rokas A (2023) A genome-informed higher rank classification of the biotechnologically important fungal subphylum *Saccharomycotina*. Studies in Mycology 105(1):1–22. doi:10.3114/sim.2023.105.01
47. Mühlhausen S, Schmitt HD, Pan KT, Plessmann U, Urlaub H, Hurst LD, Kollmar M (2018) Endogenous Stochastic Decoding of the CUG Codon by Competing Ser- and Leu-tRNAs in Ascoidea asiatica. Current Biology 28(13):2046–2057.e5. doi:10.1016/j.cub.2018.04.085
48. Žihala D, Salamonová J, Eliáš M (2020) Evolution of the genetic code in the mitochondria of Labyrinthulea (Stramenopiles). Molecular Phylogenetics and Evolution 152. doi:10.1016/j.ympev.2020.106908
49. Noutahi E, Calderon V, Blanchette M, El-Mabrouk N, Lang BF (2019) Rapid genetic code evolution in green algal mitochondrial genomes. Molecular Biology and Evolution 36(4):766–783. doi:10.1093/molbev/msz016
50. Zihala D, Eliáš M, Katz LA (2019) Evolution and Unprecedented Variants of the Mitochondrial Genetic Code in a Lineage of Green Algae. Genome Biology and Evolution 11(10):2992–3007. doi:10.1093/gbe/evz210
51. Noutahi E, Calderon V, Blanchette M, Lang FB, El-Mabrouk N (2017) CoreTracker: Accurate codon reassignment prediction, applied to mitochondrial genomes. Bioinformatics (Oxford, England) 33(21):3331–3339. doi:10.1093/bioinformatics/btx421
52. Laforest M (1997) Mitochondrial tRNAs in the lower fungus Spizellomyces punctatus: tRNA editing and UAG 'stop' codons recognized as leucine. Nucleic Acids Research 25(3):626–632. doi:10.1093/nar/25.3.626
53. Telford MJ, Herniou EA, Russell RB, Littlewood DTJ (2000) Changes in mitochondrial genetic codes as phylogenetic characters: Two examples from the flatworms. Proceedings of the National Academy of Sciences 97(21):11359–11364. doi:10.1073/pnas.97.21.11359
54. Li Y, Kocot KM, Tassia MG, Cannon JT, Bernt M, Halanych KM (2019) Mitogenomics Reveals a Novel Genetic Code in Hemichordata. Genome Biology and Evolution 11(1):29–40. doi:10.1093/gbe/evy254
55. Perseke M, Bernt M, Stadler PF, Schlegel M, Bernhard D (2011) The enigmatic mitochondrial genome of Rhabdopleura compacta (Pterobranchia) reveals insights into selection of an efficient tRNA system and supports mono-

phyly of Ambulacraria. BMC Evolutionary Biology 11(1). doi:10.1186/1471-2148-11-134
56. Abascal F, Posada D, Knight RD, Zardoya R (2006) Parallel evolution of the genetic code in arthropod mitochondrial genomes. PLoS Biology 4(5):711–718. doi:10.1371/journal.pbio.0040127
57. Abascal F, Posada D, Zardoya R (2012) The evolution of the mitochondrial genetic code in arthropods revisited. Mitochondrial DNA 23(2):84–91. doi:10.3109/19401736.2011.653801
58. Lavrov DV, Pett W, Voigt O, Wörheide G, Forget L, Lang BF, Kayal E (2013) Mitochondrial DNA of Clathrina clathrus (Calcarea, Calcinea): Six Linear Chromosomes, Fragmented rRNAs, tRNA Editing, and a Novel Genetic Code. Molecular Biology and Evolution 30(4):865–880. doi:10.1093/molbev/mss274
59. Jacob JE, Vanholme B, Van Leeuwen T, Gheysen G (2009) A unique genetic code change in the mitochondrial genome of the parasitic nematode Radopholus similis. BMC Research Notes 2. doi:10.1186/1756-0500-2-192
60. Lind C, Sund J, Åqvist J (2013) Codon-reading specificities of mitochondrial release factors and translation termination at non-standard stop codons. Nature Communications 4. doi:10.1038/ncomms3940
61. Haen KM, Lang BF, Pomponi SA, Lavrov DV (2007) Glass Sponges and Bilaterian Animals Share Derived Mitochondrial Genomic Features: A Common Ancestry or Parallel Evolution? Molecular Biology and Evolution 24(7):1518–1527. doi:10.1093/molbev/msm070
62. Wood TC (2006) The Current Status of Baraminology. Creation Research Society Quarterly 43:149–158
63. Scherer S (1994) Basic Types of Life. In: The Proceedings of the International Conference on Creationism, volume 3, 467–484
64. Scherer S, ed (1993) Typen des Lebens. Studium integrale. Pascal (Berlin)
65. Behe MJ (2019) Darwin Devolves: The New Science about DNA That Challenges Evolution. HarperOne, an imprint of HarperCollins Publishers (New York, NY), first edition
66. Mayr E (1969) Principles of Systematic Zoology. McGraw-Hill (United States of America)
67. Garen A, Siddiqi O (1962) Suppression of Mutations in the Alkaline Phosphatase Structural Cistron of E. Coli*. Proceedings of the National Academy of Sciences of the United States of America 48(7):1121–1127. doi:10.1073/pnas.48.7.1121
68. Weigert MG, Garen A (1965) Amino acid substitutions resulting from suppression of nonsense mutations. Journal of Molecular Biology 12(2):448–455. doi:10.1016/S0022-2836(65)80267-4
69. Eggertsson' G, Soll D (1988) Transfer Ribonucleic Acid-Mediated Suppression of Termination Codons in Escherichia coli. MICROBIOL. REV. 52. doi:10.1128/mr.52.3.354-374.1988
70. Biddle W, Schmitt MA, Fisk JD (2015) Evaluating Sense Codon Reassignment with a Simple Fluorescence Screen. Biochemistry 54(50):7355–7364. doi:10.1021/acs.biochem.5b00870
71. Döring V, Marlière P (1998) Reassigning Cysteine in the Genetic Code of Escherichia coli. Genetics 150(2):543–551. doi:10.1093/genetics/150.2.543
72. Saks ME, Sampson JR, Abelson J (1998) Evolution of a Transfer RNA Gene Through a Point Mutation in the Anticodon. Science 279(5357):1665–1670. doi:10.1126/science.279.5357.1665
73. Johnson DBF, Wang C, Xu J, Schultz MD, Schmitz RJ, Ecker JR, Wang L (2012) Release Factor One Is Nonessential in *Escherichia coli*. ACS Chemical Biology 7(8):1337–1344. doi:10.1021/cb300229q
74. Watanabe K, Yokobori Si (2011) tRNA Modification and Genetic Code Variations in Animal Mitochondria. Journal of Nucleic Acids 2011:1–12. doi:10.4061/2011/623095
75. Eliseev B, Kryuchkova P, Alkalaeva E, Frolova L (2011) A single amino acid change of translation termination factor eRF1 switches between bipotent and omnipotent stop-codon specificity. Nucleic Acids Research 39(2):599–608. doi:10.1093/nar/gkq759
76. Nicholson D, Salamina M, Panek J, Helena-Bueno K, Brown CR, Hirt RP, Ranson NA, Melnikov SV (2022) Adaptation to genome decay in the structure of the smallest eukaryotic ribosome. Nature Communications 13(1):591. doi:10.1038/s41467-022-28281-0
77. Plaza Oñate F, Roume H, Almeida M (2022) Recovery of Metagenome-Assembled Genomes from a Human Fecal Sample with Pacific Biosciences High-Fidelity Sequencing. Microbiology Resource Announcements 11(6):e00250–22. doi:10.1128/mra.00250-22
78. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, Lawley TD, Finn RD (2019) A new genomic blueprint of the human gut microbiota. Nature 568(7753):499–504. doi:10.1038/s41586-019-0965-1
79. Wang W, Hu H, Zijlstra RT, Zheng J, Gänzle MG (2019) Metagenomic reconstructions of gut microbial metabolism in weanling pigs. Microbiome 7(1):48. doi:10.1186/s40168-019-0662-1
80. Campanaro S, Treu L, Rodriguez-R LM, Kovalovszki A, Ziels RM, Maus I, Zhu X, Kougias PG, Basile A, Luo G, Schlüter A, Konstantinidis KT, Angelidaki I (2020) New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. Biotechnology for Biofuels 13(1):25. doi:10.1186/s13068-020-01679-y
81. Moreira D, Zivanovic Y, López-Archilla AI, Iniesto M, López-García P (2021) Reductive evolution and unique predatory mode in the CPR bacterium Vampirococcus lugosii. Nature Communications 12(1):2454. doi:10.1038/s41467-021-22762-4
82. Yakimov MM, Merkel AY, Gaisin VA, Pilhofer M, Messina E, Hallsworth JE, Klyukina AA, Tikhonova EN, Gorlenko VM (2022) Cultivation of a vampire: ' *Candidatus* Absconditicoccus praedator'. Environmental Microbiology 24(1):30–49. doi:10.1111/1462-2920.15823
83. Sieber CMK, Paul BG, Castelle CJ, Hu P, Tringe SG, Valentine DL, Andersen GL, Banfield JF (2019) Unusual Metabolism and Hypervariation in the Genome of a Gracilibacterium (BD1-5) from an Oil-Degrading Community. mBio 10(6):e02128–19. doi:10.1128/mBio.02128-19
84. Letcher PM, Lopez S, Schmieder R, Lee PA, Behnke C, Powell MJ, McBride RC (2013) Characterization of Amoeboaphelidium protococcarum, an algal parasite new to the Cryptomycota isolated from an outdoor algal pond used for the production of biofuel. PLoS ONE 8(2):e56232. doi:10.1371/journal.pone.0056232
85. Yubuki N, Zadrobílková E, Čepička I (2017) Ultrastructure and Molecular Phylogeny of *Iotanema spirale* gen. nov. et sp. nov., a New Lineage of Endobiotic Fornicata with Strikingly Simplified Ultrastructure. Journal of Eukaryotic Microbiology 64(4):422–433. doi:10.1111/jeu.12376
86. Siddall ME, Brooks DR, Desser SS (1993) Phylogeny and the reversibility of parasitism. Evolution 47(1):308–313. doi:10.1111/j.1558-5646.1993.tb01219.x
87. Leander BS, Keeling PJ (2004) Symbiotic innovation in the Oxymonad Streblomastix strix. The Journal of Eukaryotic Microbiology 51(3):291–300. doi:10.1111/j.1550-7408.2004.tb00569.x
88. Irvine DE, Guiry MD, Tittley I, Russell G (1975) New and interesting marine algae from the Shetland Isles. British Phycological Journal 10(1):57–71. doi:10.1080/00071617500650061
89. Loewe L (2006) Quantifying the genomic decay paradox due to Muller's ratchet in human mitochondrial DNA. Genetical Research 87(2):133–159. doi:10.1017/S0016672306008123

90. Lynch M (1996) Mutation accumulation in transfer RNAs: Molecular evidence for Muller's ratchet in mitochondrial genomes. Molecular Biology and Evolution 13(1):209–220. doi:10.1093/oxfordjournals.molbev.a025557
91. Neiman M, Taylor DR (2009) The causes of mutation accumulation in mitochondrial genomes. Proceedings of the Royal Society B: Biological Sciences 276(1660):1201–1209. doi:10.1098/rspb.2008.1758
92. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, Karpinets T, Lund O, Kora G, Wassenaar T, Poudel S, Ussery DW (2015) Insights from 20 years of bacterial genome sequencing. Functional & Integrative Genomics 15(2):141–161. doi:10.1007/s10142-015-0433-4
93. Pray L (2008) Eukaryotic Genome Complexity — Learn Science at Scitable. Nature Education 1(1):96
94. Crick F (1963) The recent excitement in the coding problem. Progress in Nucleic Acid Research and Molecular Biology 1:163–217. doi:10.1016/S0079-6603(08)60642-X
95. Hinegardner RT, Engelberg J (1963) Rationale for a Universal Genetic Code. Science 142(3595):1083–1085. doi:10.1126/science.142.3595.1083
96. Crick F, Orgel L (1973) Directed panspermia. Icarus 19(3):341–346. doi:10.1016/0019-1035(73)90110-3
97. Jukes TH, Osawa S (1997) Point counter point. Journal of Molecular Evolution 45(1):1–3. doi:10.1007/PL00006192
98. Osawa S, Jukes T (1995) On codon reassignment. Journal of Molecular Evolution 41(2). doi:10.1007/BF00170679
99. Schultz DW, Yarus M (1996) On malleability in the genetic code. Journal of Molecular Evolution 42(5):597–601. doi:10.1007/BF02352290