CellPress

Current Biology

Review

Natural and artificial variations of the standard genetic code

Julius Lukeš[1,7], Marek Eliáš[2,7,*], Ambar Kachale[1,3,4], Peter T.S. van der Gulik[5], and Dave Speijer[6,*]
[1]Institute of Parasitology, Biology Centre, Czech Academy of Sciences, and Faculty of Science, University of South Bohemia, České Budějovice (Budweis) 37005, Czech Republic
[2]Department of Biology and Ecology, Faculty of Science, University of Ostrava, Ostrava 71000, Czech Republic
[3]Department of Cellular and Molecular Biology, Harvard University, Cambridge, MA 02138, USA
[4]Current address: Department of Biochemistry, University of Cambridge, Cambridge CB2 1QW, UK
[5]Centrum Wiskunde and Informatica, Amsterdam 1098 XG, The Netherlands
[6]Medical Biochemistry, Amsterdam UMC, University of Amsterdam, Amsterdam 1105 AZ, The Netherlands
[7]These authors contributed equally
*Correspondence: d.speijer@amsterdamumc.nl (D.S.), marekelias1978@gmail.com (M.E.)
https://doi.org/10.1016/j.cub.2025.09.071

SUMMARY

The nearly universal standard genetic code is one of the strongest indications for the common ancestry of all extant life. Yet new code variants representing different combinations of codon reassignments have been and continue to be discovered with regular frequency. Recently, many man-made codes have been generated that direct the incorporation of unnatural amino acids, allowing for biocontainment and viral resistance. However, evolutionary and synthetic biologists often seem to be unaware of the developments in the other camp's field. Here we attempt to bridge this gulf and provide an updated overview of different codon reassignments and genetic code variants reported to occur naturally in organisms and their organelles, both now containing over 50 examples. We review the highly expanded range of departures from the standard genetic code, highlighting previously unanticipated code forms and their molecular underpinnings. To account for cases when a codon has a different meaning depending on its context, such as in variants with no dedicated termination codons discovered over the past decade in protists, we introduce the concept of codon homonymy. Considering this new appreciation for the prevalence of genetic code diversity, we also revisit the questions of how and why genetic codes get altered in evolution. Finally, we summarize the current status of artificially redesigned genetic codes, which are increasingly deviating from natural code alterations, opening up completely novel translational possibilities.

Introduction

The genetic code is as essential to understanding biology as the periodic table is to understanding chemistry. In chemistry, 94 known elements occur naturally, whereas 25 have been generated by humans within a bit over a century[1]. For the genetic code, things are very different: one version, the canonical or standard genetic code, predominates, yet over 50 natural alternatives have been identified, although most are rare (Table S1), and the rate of discovery is plateauing, despite the enormous growth of relevant datasets. However, the number of alternative man-made genetic codes is growing rapidly[2–7]. Presumably, some novel, naturally occurring alternative genetic codes remain to be discovered, but they will be swamped by the number of artificial versions generated. Indeed, synthetic biology allows us to generate almost endless combinations of the genetic code variations with different aims, such as providing barriers to viral infections, prevention of unwanted horizontal gene transfer from genetically modified organisms, and incorporation of non-standard amino acids[8–11].

The genetic code is a set of rules used by cells to translate information encoded by mRNAs into the amino acid sequence of proteins. The complete elucidation of the standard genetic code belongs to the key discoveries of the 20th century. It is now textbook knowledge that the coding sequence is interpreted during translation as a series of non-overlapping nucleotide triplets, called codons, each designating either one of the 20 common proteinogenic amino acids, or serving as translation termination signals (UAA, UAG, UGA) that are recognized by release factors instead of tRNAs[6,12]. Four different nucleotides give 64 possible codons, making the standard genetic code necessarily degenerate, most amino acids being specified by multiple (two to six) synonymous codons (Figure 1). Of course, translation guided by the genetic code refers to only a single part of the process by which genes are decoded. Other aspects, including the maturation of transcripts by cis- and/or trans-splicing, RNA editing, and 3'-end processing and polyadenylation[13,14], can influence the coding region before it is interpreted during translation. Of note, translation initiation (that is, start codon selection) is also governed by rules independent of the genetic code sensu stricto[15–17]. Occasionally, a codon other than the 'standard' AUG is used to initiate translation in both prokaryotes and eukaryotes[18,19]. In such instances, identical initiator tRNAs decode different initiation codons, ignoring 'normal' rules of codon–tRNA pairing that operate during elongation[18,19].

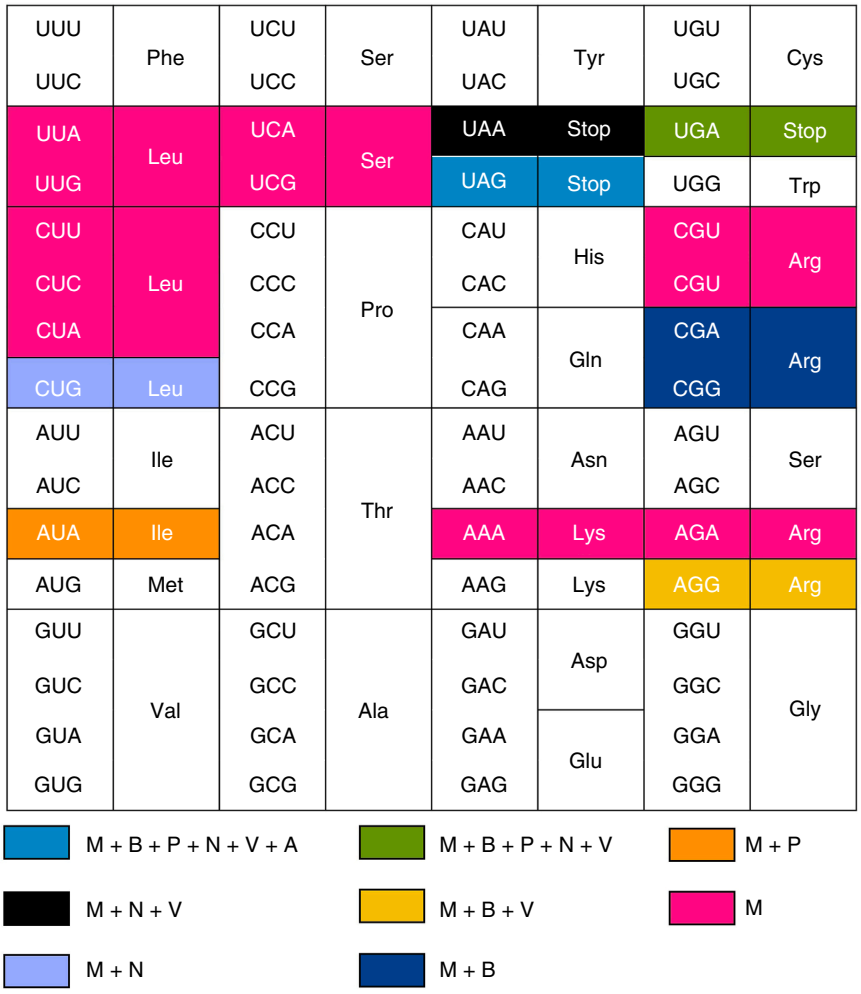


Figure 1. Standard genetic code table.
The standard 64 triplets are shown. Of these, the 19 triplet combinations that are known to have been naturally reassigned anywhere in the tree of life are indicated in colours, reflecting where they have been found. Of note, UAG reassignment in Bacteria is presently restricted only to the polysemous decoding as Pyl parallel to the standard translation termination; see main text. For simplicity, the homonymous UGA decoding as Sec, which we consider to be part of the standard genetic code (Box 1), is not indicated.

Molecular implementation of the standard genetic code is understood in great detail because of its virtual universality, straightforward logic, and mechanistic elegance: tRNAs carrying a specific amino acid have anticodons that base pair with the right set of codons. The identity of the amino acid carried by a particular tRNA is defined predominantly by dedicated aminoacyl-tRNA synthetases (herein referred to as synthetases), which charge one or several tRNAs with a specific amino acid[20], though an alternative route exists in some cases (Box 1). The interaction between the tRNA's anticodon and mRNA's codons is commonly extended by wobble base-pairing of the third codon/first anticodon position, which allows a tRNA to recognize and bind to multiple (synonymous) codons encoding the same amino acid. Consequently, the number of distinct tRNAs employed by a given translation system is usually substantially lower than the number of codons, but often not near the theoretical minimum, because of the existence of so-called isoacceptor tRNA species carrying identical amino acids, but pairing with different codons. In eukaryotes we typically find ~45 isoacceptor tRNAs[21]. The copy number of different tRNA genes varies as well, with single tRNA-species genes predominating in prokaryotes and a steady increase in copy numbers, correlating with genome size, in eukaryotes[22]. A rich assortment of base modifications further impacts the specificity of codon–anticodon pairing, either extending or restricting the possible combinations allowed[23,24]. Translation of the nucleotide sequence is accomplished by ribosomes and suites of translation factors, linking amino acids in the order specified by mRNAs[25]. Amino acid-charged tRNAs consecutively match single codons in the ribosomal A-site, followed by peptide-bond formation in its P-site, and exit of the de-acetylated tRNA via the E-site. Termination of translation is mediated by release factors, binding to stop codons and subsequently liberating the completed polypeptide from the tRNA that paired with the preceding codon.

The operation of the genetic code described above is, as usual in biology, an idealization of the real state of affairs. Complicating the picture is inherent molecular noise (reflecting the stochastic nature of molecular interactions) that limits decoding fidelity. Thus, with a certain frequency, unintended amino acids are introduced into nascent polypeptides, translation terminates prematurely, and/or termination codons are read through[26–29]. The frequency of such translation errors (the mistranslation rate) may fluctuate, for example, during stress or disease[26,30], but also varies between different organisms. For example, synthetases are normally endowed with editing domains that selectively remove a non-cognate aminoacyl-adenylate (pre-transfer editing) or a non-cognate amino acid from tRNA (post-transfer editing)[31]. However, certain endobiotic or parasitic bacteria and eukaryotes exhibit editing-domain degeneration, resulting in increased tRNA mischarging and mistranslation[32,33] (Box 2). Naturally occurring human tRNA gene variants can cause mistranslation as well[34]. It is debated whether the increased proteome diversity resulting from such mistranslation is adaptive[35] or whether, in general, it merely reflects limitations of selection in imposing higher translational fidelity[36,37].

Box 1. The standard genetic code and Theseus's paradox.

The common definition of the standard genetic code is that it is the most widespread genetic-code variant tabulated as shown in Figure 1. However, this representation ignores the commonly occurring UGA homonymy, that is, the alternative context-dependent decoding of the codon as Sec (selenocysteine). Conservation and broad phylogenetic distribution of components allowing selenocysteine utilization via UGA recoding attest to its presence already in the last universal common ancestor (LUCA). We thus argue that the SECIS element–SelB-dependent UGA decoding forms part of the standard genetic code. From this perspective, organisms (and organelles) that secondarily lost this feature exhibit a non-standard genetic-code variant. But what about UAG as the (polysemous) stop and pyrrolysine (Pyl) codon? Here the situation is more complicated. The phenomenon occurs in only some archaeal and bacterial lineages and seems easily explained by a later origin and subsequent horizontal gene transfer-mediated dissemination, were it not for phylogenetic evidence of the Pyl-tRNA synthetase being as ancient as other aminoacyl-tRNA synthetases[195]. Maybe the Pyl incorporation machinery and genes relying on it were present in the LUCA's ancestors, surviving in some descendant lineages but not in the LUCA itself. Such 'sister' lineages could then have donated the Pyl machinery to some of the LUCA's descendants before going extinct[195]. According to this hypothetical scenario, Pyl encoding would be ancient but not part of the standard genetic code.
But there are deeper problems with the standard genetic code concept: although codon meaning generally remains constant, underlying components and mechanisms have continued evolving since the LUCA. Release factors in bacteria compared with archaea and eukaryotes are non-homologous, despite their roughly similar structure (mimicking a tRNA shape) and sharing a conserved motif involved in the liberation of the completed polypeptide chain from the final tRNA[52,53]. In eukaryotes and archaea, a single release factor operates whereas, counterintuitively, bacteria and organelles employ two unrelated release factors. Which of these two alternative mechanisms is the ancestral one remains unclear; it was even speculated that the LUCA utilized a tRNA-like RNA terminator, which was later independently replaced by the proteinaceous release factors[53].
The mechanisms of producing the same aminoacyl-tRNA species may also vary. For glutamine, asparagine, and cysteine, the respective aminoacyl-tRNAs are synthesized either directly by dedicated synthetases or indirectly by charging the tRNA with a precursor amino acid (glutamate, aspartate, and phosphoserine, respectively) followed by a modification of the aminoacyl moiety by a dedicated enzyme[196,197]. Whereas the production of these three aminoacyl-tRNAs occurs via the direct route for use during cytoplasmic translation in eukaryotes, prokaryotes and organelles vary in their deployment of direct and indirect pathways, indicating a complex evolutionary history involving horizontal gene transfer and replacements. The synthetase dedicated to the production of Gln-tRNA (GlnRS) seems to be a eukaryotic innovation that was subsequently introduced into several bacterial lineages, including γ-proteobacteria, replacing the ancestral indirect mechanism[197,198]. It is striking to realize that translation in *E. coli*, the prototypical bacterium, relies on a component invented by eukaryotes! Alternatives exist also when direct pathways of aminoacyl-tRNA synthesis are compared. For example, there are two unrelated synthetase forms dedicated to lysyl-tRNA synthesis (LysRS) that evolved independently within the two major aminoacyl-tRNA synthetase groups (classes I and II). The class II LysRS dominates across the tree of life, whereas the class I LysRS occurs only sporadically in various bacterial and archaeal lineages[153].
There also exists a wide variety of tRNA modifications, a suite of alterations important for tRNA folding, stability, aminoacylation, as well as the decoding function itself[24]. A perfect case in point is the different modifications critical for proper decoding of AUA as isoleucine (specifically, those that allow discrimination of AUA from AUG, which encodes methionine). In bacteria, their descendant organelles, and archaea, the $tRNA^{Ile}$ employed for this task differs from the tRNA decoding the two other isoleucine codons (AUU, AUC). Specific cytidine derivatives at the first (wobble) anticodon position (lysidine in bacteria and organelles, agmatidine in archaea) allow pairing with A but not G at the third codon position[199]. However, exceptions exist in Bacteria, such as in the parasite *Mycoplasma mobile*, which lacks the conventional $tRNA^{Ile}_{CAU}$ and its modifying enzyme, instead possessing a unique $tRNA^{Ile}_{UAU}$ with an unmodified anticodon that reads AUA, while decoding of AUG by this tRNA is somehow avoided by its ribosome[200]. In eukaryotic translation, AUA decoding primarily relies on a tRNA with inosine at the first anticodon position (resulting from adenosine deamination), which also decodes AUU and AUC. In addition, a second tRNA with the anticodon ΨAΨ (Ψ denoting pseudouridine introduced by enzymatic conversion of uridine), specifically decoding AUA, is also present. The diverse ways of AUA decoding across the tree of extant life might indicate that this codon was not used by the LUCA, only to be captured by isoleucine independently in the bacterial and archaeal stem lineages later[201]. Why and how the eukaryotic mechanism of AUA decoding evolved remains an open question.
Apart from these and other major differences in the implementation of the standard genetic code, there is a rampant, albeit less conspicuous, turnover of translation apparatus components across the tree of life. For example, tRNAs get replaced by functionally equivalent but completely unrelated new ones, including mutated duplicates of tRNAs originally with different anticodon and amino acid specificities (alloacceptor tRNAs), a process referred to as 'tRNA gene recruitment' or 'tRNA remolding'[129,202,203]. Most mitochondrial genomes lost genes for many, sometimes all, tRNAs, and the missing ones imported from the cytosol are of eukaryotic origin (that is, they did not result from gene translocations of original mitochondrial genes to nuclear genomes)[204,205]. Horizontal gene transfer has been a major driver of aminoacyl-tRNA synthetase evolution, causing pervasive xenologous gene displacement, with donors among cellular organisms of varying phylogenetic distance from the recipients as well as viruses[206]. Certain green algae utilize the UGA stop codon in their mitochondria despite lacking the cognate mitochondrial release factor (mtRF2a), and dual targeting of the cyanobacteria-derived plastidial homolog (pRF2) has been proposed as a mechanism ensuring proper UGA decoding in these cases[78]. Altogether, the standard genetic code concept reminds us of the mythical Ship of Theseus: having replaced all the planks, is it still the same ship?

Box 2. Glossary.

Ambiguous decoding — In the literature the term is often uncritically used to refer to all instances in which particular codons are translated with two or more differing outcomes at a level significantly above what could be expected from the intrinsic limits of translational reliability. However, underlying mechanisms and outcome distributions can be completely different; that is to say, the term 'ambiguous' is itself used ambiguously to refer to incomparable phenomena. We distinguish between codon polysemy and codon homonymy below.

Orthogonality — This concept originally referred to the ability of a tRNA to interact exclusively with its corresponding synthetase while avoiding cross-reactions with additional types of tRNAs and synthetases present in a given organism. This state of affairs has to be retained when expanding the genetic code[184]. With the explosive development of artificial coding it is now used to describe groups of molecular components (synthetases, tRNAs, mRNAs and ribosomes) restricted to only recognize each other during translation, thus avoiding or minimizing cross-reactions with the endogenous translation machinery. Even more broadly, in synthetic biology, for example, when introducing artificial non-canonical amino acids in a living cell, it is used to refer to minimizing any kind of unwanted interaction with host machinery[207].

Codon homonymy — We introduce this term to refer to instances where a codon can have multiple meanings that are distinguished by a given translation system based on the specific context in which the codon occurs. UGA as a termination codon (by default) or as a selenocysteine codon (in the case of mRNAs with SECIS structures) is the classical example of codon homonymy, whereas the systemic (i.e., pangenomic) deployment of certain codons as sense ones (when in-frame) or as termination ones (when at the 3' end of the coding sequence) occurring in multiple protist lineages including various ciliates or the trypanosomatid *Blastocrithidia* is an even more spectacular manifestation of the phenomenon.

Codon polysemy — This refers to instances where a codon has multiple meanings and its translational outcome seems to depend on chance. This does not mean the system will always end up in 50:50 distributions but rather that individual codons are not biased with regard to the outcome. The term was coined first in 1997, though real examples were found only later, for example, in the yeast *Ascoidea asiatica*.

Codon reassignment — This term applies when, in principle, all instances of a codon have obtained a new meaning in a cell or an organelle. Three categories can be recognized: stop-to-sense, sense-to-sense, and sense-to-stop codon reassignments. Codons systemically used as ribosomal frameshifting signals, exemplified by in-frame UAA and UAG codons in the ciliate genus *Euplotes* (sensu lato), may be conceptualized to represent a codon reassignment category of its own.

Mischarging — This term denotes the production of a tRNA charged with an amino acyl residue not cognate to the tRNA considering the organism's genetic code. Of note, its connotation ('mischarging' leads to the incorporation of the 'wrong' amino acid) obscures the many instances where this mechanism is actually part of an appropriate response. An interesting possibility: reactive oxygen species (ROS)-mediated cell damage may trigger significant protective methionine misacylation of non-Met tRNAs, resulting in a general increase in methionine incorporation, buffering the oxidative damage[35,208].

Recoding — This refers to instances in which some triplets are translated differently from the meaning prescribed by the genetic code employed by the given translation system. Mechanistically, recoding depends on an interplay between the transcript context of the triplet in question and the components of the translational system, as well as regulatory factors external to it, such as availability of particular metabolites. A codon at a given position may be translated as recoded by default (for example, an UGA codon upstream of a SECIS element) or its translation may vary between a standard and recoded one depending on such specific physiological clues.

Another category of irregularities complicating the simple logic described above includes phenomena exemplified by the phylogenetically widespread co-translational incorporation of selenocysteine (Sec). The mechanism relies on a local redefinition of default codon meaning, typically of UGA (stop codon), dependent on a specific secondary structure in the mRNA (the SECIS element) and a specific translation elongation factor controlling the insertion of $tRNA^{Sec}$ into the ribosomal A-site[38]. Processes including programmed ribosomal frameshifting, translational bypassing, or regulated stop-codon readthrough mediated by specific molecular signals are further examples where the interpretation of a particular region (often a single codon) of the mRNA departs from the standard rules of the organism's genetic code. Here, specific molecular factors or signals impinge upon the process of translation[39]. Such phenomena can conveniently be grouped under the rubric of 'recoding'[40,41]. However, we acknowledge that this term is used more fluidly, frequently referring to a conceptually different situation in which particular codons systemically (that is, across all genes in a given species or taxonomic clade) diverge from the standard meaning. Such changes, more appropriately called codon 'reassignments', are what distinguish different genetic code variants and will be extensively discussed below.

Before embarking on this topic, we feel further clarification and expansion of the existing vocabulary used for describing different modes of codon interpretation are in order (see also Box 2 for a Glossary of key terms). Certain recoding and reassignment cases are often described as 'ambiguous decoding', but as argued elsewhere[42], this term is vague and describes very different phenomena in different contexts. In one scenario, a particular codon is systemically interpreted in two (or hypothetically more) parallel ways, regardless of the gene and the position of the codon in the coding sequence. The term 'polysemous codon' was once proposed for such a configuration[43], and

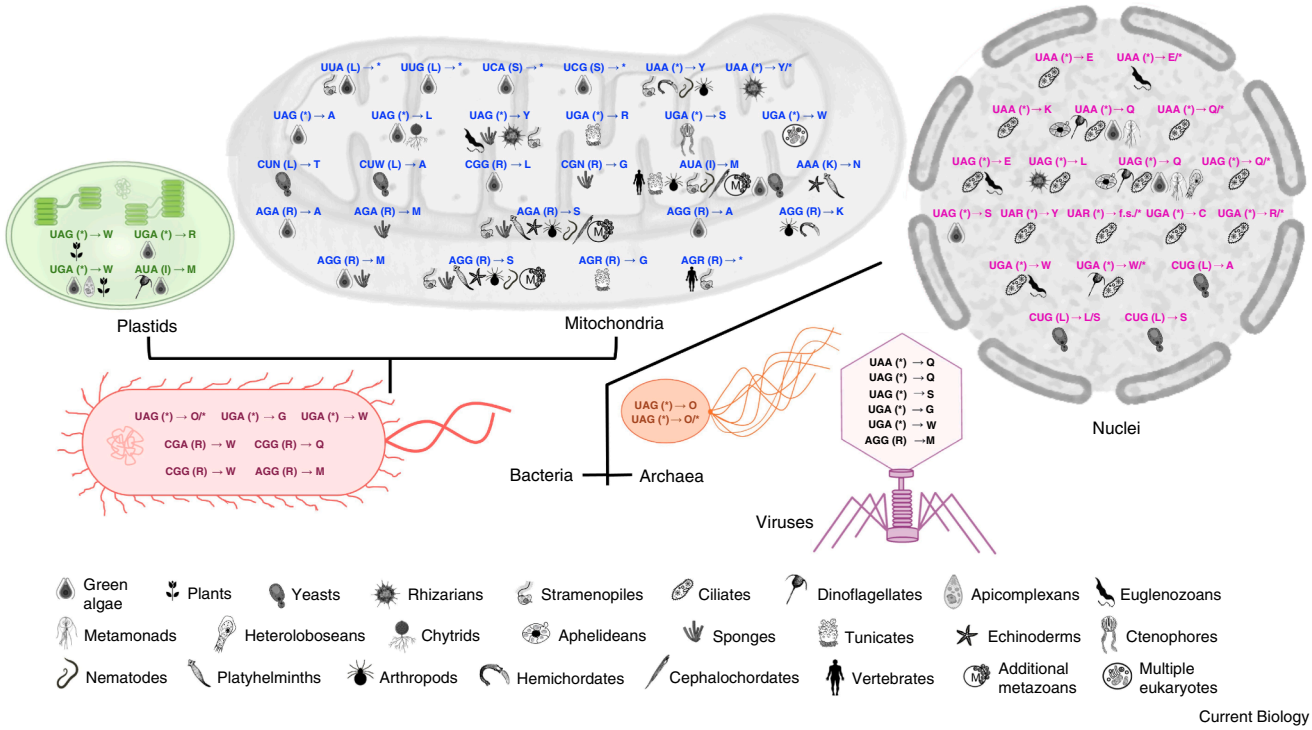


Figure 2. Distribution of codon reassignments across all domains of life, organelles and viruses.
All well-documented cases are shown in the form of the affected codon or group of related codons followed by the standard one-letter code of the amino acid in the standard genetic code (bracketed) and the amino acid upon reassignment. IUPIAC base codes: N = A+C+G+U; R = A+G; W = A+U. The letter O denotes Pyl; asterisks (*) indicate termination codons; the abbreviation f.s. stands for frameshifting signal; and the slash symbol (/) separates two alternative meanings of a codon in the case of codon homonymy or polysemy. The homonymous meaning of UGA as a Sec codon, as well as the multiple codons serving as frameshifting signals in the *Perkinsus* mitochondrion (see main text) are omitted for simplicity. Icons schematically depict groups of organisms where the reassignments were detected (see graphical legend for details); the reassignments are not necessarily present in all members of a given group. References for the taxonomic occurrence of the codon reassignments are provided in Tables S1 and S2. See also Conclusions and outlook for the very recently reported codon reassignment cases that could not be included in the figure.

with recent reports of what qualifies as bona fide codon polysemy[44,45], we propose that the term should be revived. A different situation occurs when codons have distinct meanings depending on their context, as exemplified by the aforementioned UGA decoding as Sec at certain positions in certain mRNAs. We propose the term 'homonymous codon' be used in such cases. Thus, the standard genetic code contains many synonymous codons (different codons with the same meaning) as well as the homonymous UGA codon (Box 1). Other, spectacular, lineage-specific cases of codon homonymy are discussed later.

The genetic code is evolutionarily malleable

How the standard genetic code arose is one of the toughest biological questions around with no generally accepted scenario available. The standard genetic code has been labelled as a 'frozen accident'[46] and considered refractory to changes due to its essentiality. However, out of the astronomical number of $>10^{84}$ alternatives, the standard genetic code is non-random and highly optimized (with the level of optimization dependent on parameters, for example error robustness, considered[47,48]), albeit not the absolute best[49,50]. Regardless, its composition seems to be highly resistant to the impact of point mutations, since mutated codons are usually synonymous or specify an amino acid with similar biochemical properties[51]. Crucially, the standard genetic code is itself subject to change. Hence, describing it as 'frozen with occasional melting' more accurately reflects reality: it can evolve under specific circumstances, showing minor alterations that are compatible with cell survival. As a result, operation of the genetic code varies substantially across the diversity of extant life, and at least two different axes of this variation can be recognized.

The less conspicuous axis reflects the range of different molecular solutions resulting in the same outcome (the same codon meaning), illustrated by the many different ways in which the standard genetic code is implemented at the molecular level in different translation systems. The differences concern details of the synthesis of particular aminoacyl-tRNAs, specific modifications of tRNAs required for codon specificity, or the number and nature of release factors. For many of these, the dividing line goes between bacteria (and their mitochondrial and plastid derivatives) and archaea/eukaryotes, indicating deep evolutionary origins and raising the question of which of the solutions (if any) were employed by the last universal common ancestor (LUCA) versus those that are replacements of original mechanisms. For example, translation termination in bacteria and organelles relies on two different (yet homologous) release factors, RF1 and RF2, recognizing UAG/UAA and UAA/UGA, respectively, whereas archaea and eukaryotes utilize a single, unrelated omnipotent aRF1/eRF1 to recognize all three stop codons[52,53].

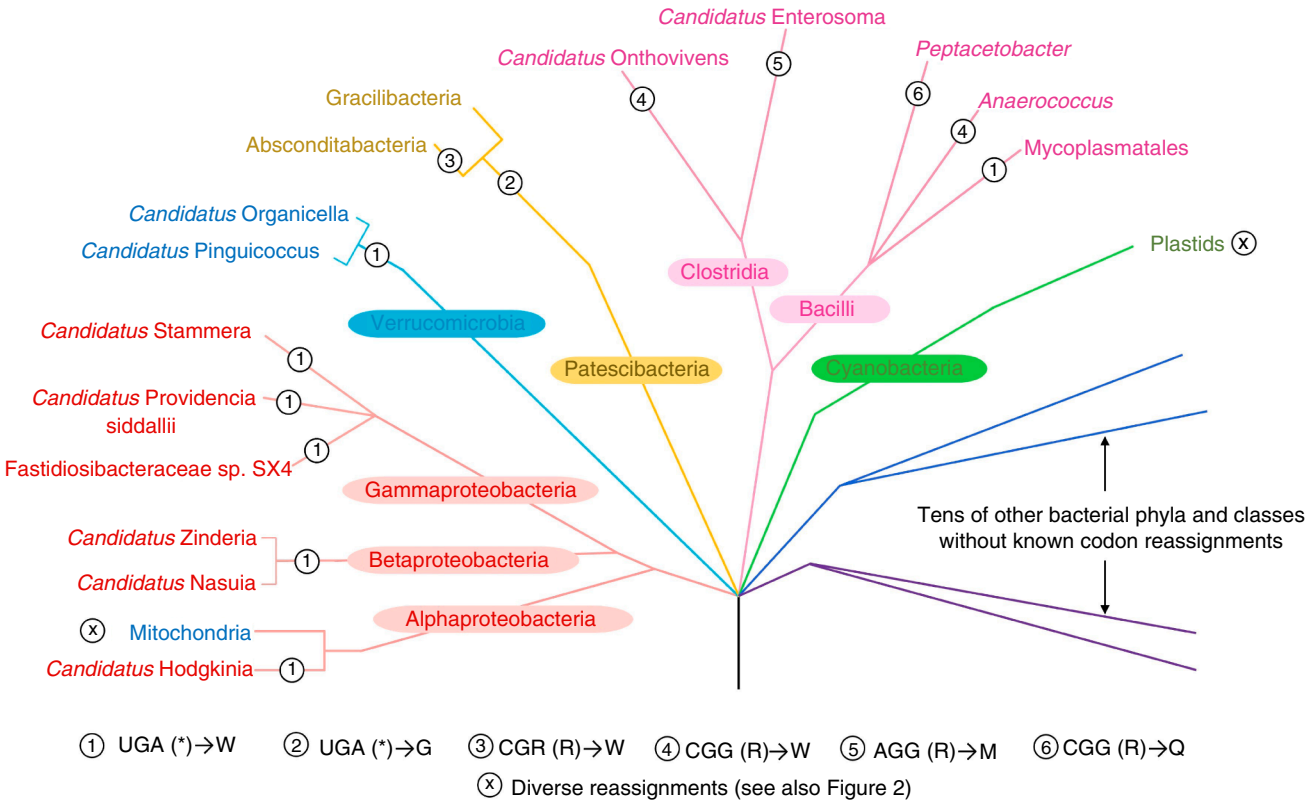


Figure 3. Phylogenetic tree of Bacteria with the distribution of known codon reassignments.
For simplicity, the phylogeny is presented with the main lineages in an unresolved polytomy stemming from a single node and species epithets are not provided unless there is evidence that different species in the same genus may utilize a different genetic code variant. Individual cases of different codon reassignments (circled numbers) are mapped onto the tree to most parsimoniously explain their occurrence in extant taxa; see legend of the tree for details on the reassignments. Cases with polysemous UAG serving as both a Pyl and termination codon are omitted from the figure. In *Candidatus* Providencia siddallii the indicated stop-to-Trp UGA reassignment concerns only one specific strain (PSAC) and the codon seems to still be used to terminate translation of a few genes (too little is known to classify this ambiguity as codon homonymy or polysemy)[109]. References for the codon reassignments displayed are provided in Tables S1 and S2.

All this variation is hidden underneath the façade of an unchanged, standard genetic code (Box 1).

The other axis along which the genetic code varies regards how different codons are interpreted: the specific amino acid (or stop signal) specified by the codon. Ever since its discovery, the genetic code was seen as extremely resistant to change because of its central role in the production of every single protein an organism generates. This brake on further evolution has been described under the heading of 'proteomic constraint'[54]. The concept of genetic code universality collapsed with the sequencing of the human mitochondrial genome[55,56]. By demonstrating that UGA is read as tryptophan and not as a termination signal, AUA is decoded as methionine instead of isoleucine, and AGA/AGG (normally specifying arginine) function as stop codons, the analysis of the human mitochondrial genome showcased all three major types of codon reassignments: stop-to-sense, sense-to-sense, and sense-to-stop.

Further reports of codon reassignments in other mitochondria, as well as in bacteria[57], eukaryotic nuclear genomes[58], and viruses[59] quickly followed, and were later complemented by the first instance of a non-canonical genetic code in a plastid[60] (Tables S1 and S2). A rich landscape of alternative genetic codes and molecular mechanisms responsible had already been uncovered before the start of the new millennium[23]. Though the pace of discovery of new genetic code variants then slowed down, it rose again during the past decade mostly as a result of systematic screening of ever more taxonomically diverse genomic and transcriptomic data. Based on an exhaustive literature survey, we compiled an up-to-date list of known codon reassignments, their combinations, and taxonomic occurrence (Figures 2–4 and Tables S1 and S2). We note that the list of genetic codes (or translation tables) provided by the NCBI (https://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi?chapter=cgencodes) is not an authoritative resource (despite frequently being considered thus), as it covers only about half of known genetic code variants, some of which are defined inaccurately (Table S1). As a result, many protein sequences maintained in GenBank (and thus various secondary sequence databases), especially for mitochondria, do not correspond to reality. Another issue: most reported cases of codon reassignments represent bioinformatic inferences based on comparative sequence analyses only and should be treated with caution.

The naturally evolved codon reassignments described so far are unevenly distributed across the code chart (Figure 1), with

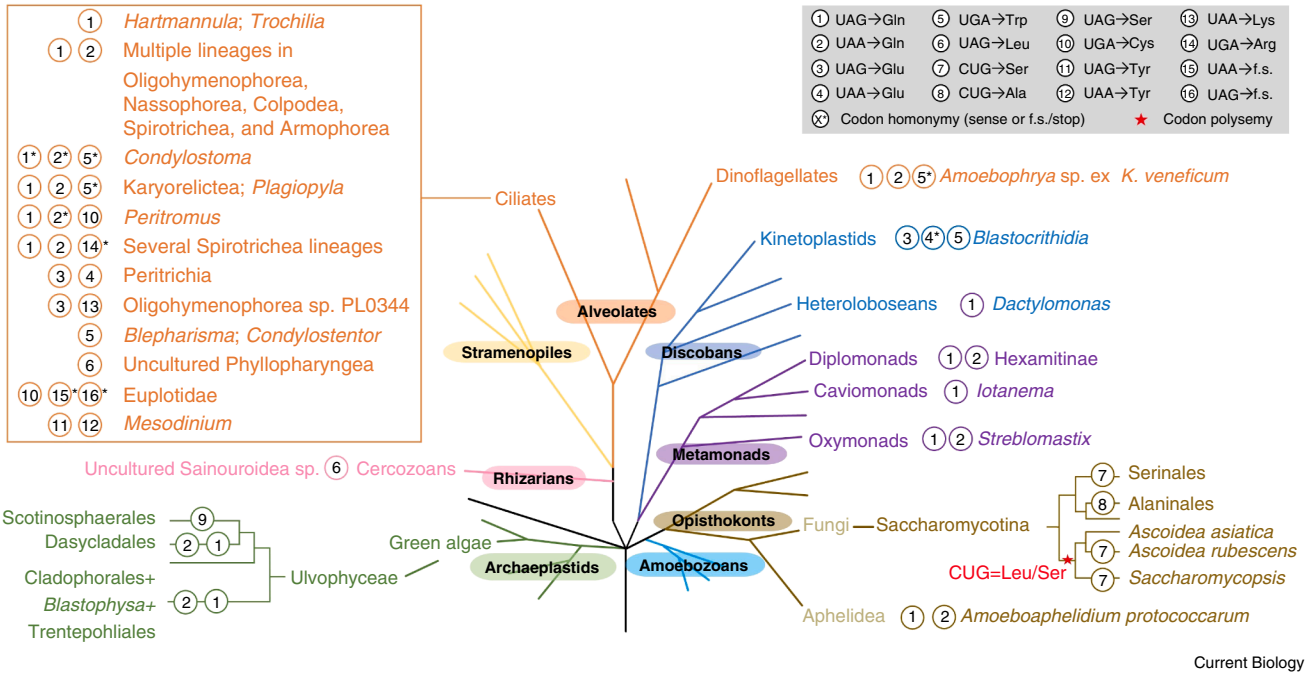


Figure 4. Phylogenetic tree of Eukarya with the distribution of the known codon reassignments.
Only reassignments (designated by circled numbers) affecting the nuclear genome are shown, mapped onto a simplified phylogeny reflecting the current consensus to most parsimoniously explain their occurrence in extant taxa; some smaller eukaryote lineages from which no codon reassignments have been reported are omitted from the scheme. For ciliates the different reassignment combinations are listed without providing their phylogenetic context, only indicating the specific ciliate taxa in which they have been reported (taxa separated by a semicolon represent independent reassignments). Cases of codon homonymy (denoted with an asterisk) and polysemy (the star symbol) are also indicated; f.s. corresponds to frameshifting signal. See the graphical legend and the main text for further details. References for the codon reassignments displayed are provided in Tables S1 and S2. A broader range of reassignment combinations with putative homonymous decoding of UAG, UAA, or UGA as sense/stop codons than is displayed in the figure has been implied to exist in ciliates by a recent study[93], but further investigations are needed to verify these results.

some codons capable of numerous alternative interpretations (the record holder currently being UAG with nine possible amino acid outcomes, including pyrrolysine) (Table S2), whereas many other codons, such as GGN, (so far) only specify the canonical amino acid (glycine) (Figure 1). However, this picture may be biased, since reassignments of stop codons are more readily recognizable due to the corresponding interruption of coding regions when they are conceptually translated with the default standard genetic code[61]. Indeed, recent targeted computational screens of prokaryotic genomes identified multiple cases of sense-to-sense reassignments[62], and it is plausible that comparing DNA sequences with mass spectrometry data will unearth more code alterations. Given current knowledge, the number of reassigned codons in a particular genetic-code variant ranges from one to six, with this high number achieved multiple times in mitochondria of different eukaryote lineages (Table S1). In Archaea the only deviation from the standard genetic code known so far concerns UAG decoding as pyrrolysine (discussed in more detail below).

Obviously, any departure from the standard genetic code resulting in codon reassignment must be underpinned by specific changes in the translation system. These changes nearly always involve a combination of gains and losses — of whole components or their particular functions[63] (Figures 5 and 6 and Box 3). The components thus affected include tRNAs (in all reassignment types) and release factors (in stop-to-sense or sense-to-stop reassignments). The role and range of synthetase changes are highly understudied: although theoretically expected to be a critical part of most codon reassignments, very few well-documented examples of specific synthetase changes associated with particular reassignments are available. Sometimes, a specific tRNA change underpinning a codon reassignment does not result from a mutation of the tRNA gene itself but depends on changes in the enzymatic machinery mediating post-transcriptional editing or modifications of the tRNA. To our knowledge, specific changes of the ribosome or elongation factors that would directly underlie naturally occurring codon reassignments have not yet been described.

The genetic-code zoo

It is beyond the scope of this review to discuss in detail all known codon reassignments and the underlying molecular mechanisms, so the reader may wish to consult other reviews[42,49,64,65] and original reports listed in Tables S1 and S2. Instead, we highlight those more recent discoveries that have most substantially changed our perception of the range of evolutionary diversification of the genetic code.

The expanded scope of sense-to-sense codon reassignments

Sense-to-sense codon reassignments were long considered to be virtually exclusive to mitochondria, the exception being the Leu-to-Ser reassignment of CUG in the nuclear genes of

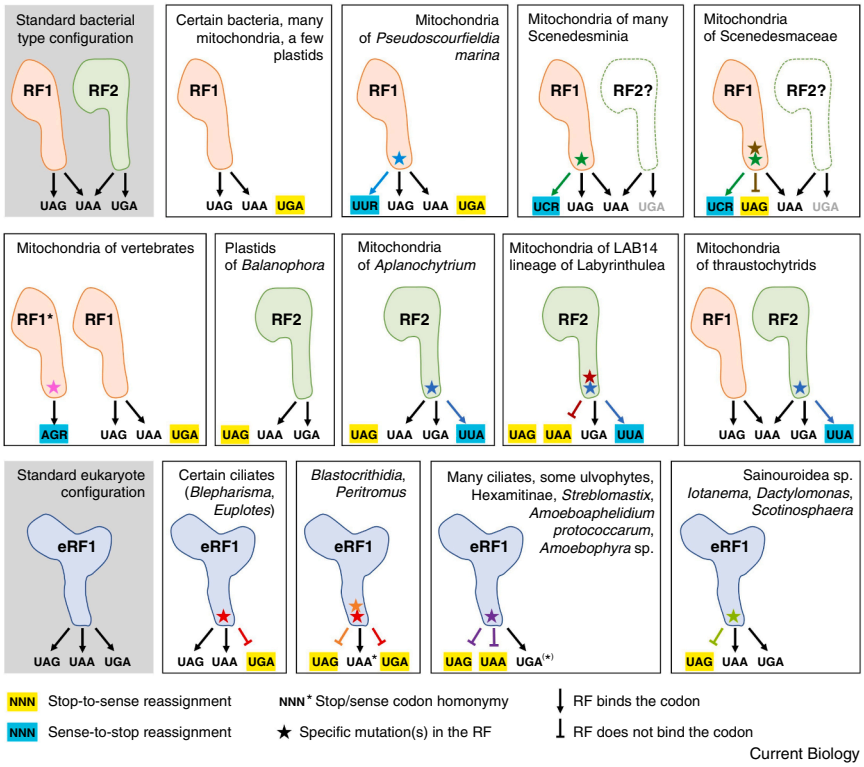


Figure 5. Changes in release factors underpinning known stop-to-sense or sense-to-stop codon reassignments.
The three different release factors, bacterial RF1 and RF2 (plus their respective organellar equivalents, mRF1a/pRF1 and mRF2a/pRF2) and the eukaryotic eRF1, are identified by name and colour. The UGA codon sporadically found in mitochondria of certain green algae (many Scenedesminia, including its subgroup Scenedesmaceae) might be recognized by dually targeted pRF2 (dashed outline), but direct evidence is lacking. The different star colours indicate the (combinations of) mutations resulting in the loss (yellow highlight) or gain (turquoise highlight) of codon binding indicated below each release factor (see main text). In vertebrate mitochondria a lineage-specific RF1 paralog (marked with an asterisk) with a specifically altered codon-binding domain mediates translation termination at the AGR codons.

Candida albicans and related yeasts[66]. Later, however, plastids of certain algae joined this class, as they exhibit the Ile-to-Met reassignment of AUA[67,68]. This parallel switch in multiple lineages of both mitochondria and plastids is an obvious manifestation of the reductive evolution affecting organelles in general, in the latter case involving the loss of a unique $tRNA^{Ile}$ and the dedicated enzyme engaged in modifying its anticodon to make it cognate to AUA (see Box 3 for details). Sense-to-sense reassignments found in plastids are perhaps less surprising than the result of a recent systematic screen of prokaryotic genomes, using the dedicated computer tool Codetta[62]. This screen revealed, for the first time, sense-to-sense codon reassignments in lineages of free-living bacteria, specifically the reassignment of some arginine codons to methionine, tryptophan or glutamine[62]. Interestingly, this set of reassignments does not overlap at all with the rich repertoire identified in mitochondria (Table S2), suggesting that different evolutionary constraints shape the genetic code evolution in bacteria and their organellar descendants. Finally, the story of the CUG codon in yeasts turned out to be much more complex than envisioned earlier (detailed below)[69,70].

Codon polysemy revisited

A few studies suggested that CUG in *Candida* and related yeasts is decoded ambiguously, with the cognate, newly evolved $tRNA^{Ser}_{CAG}$ experiencing a considerable level of mischarging by leucine. The concept of CUG as a polysemous codon (Box 2) was proposed, along with speculation that the resulting proteome heterogeneity confers selective advantages[43,71,72]. However, although initial findings supported substantial alternative CUG decoding as leucine[73], methodologically more advanced analyses later showed CUG decoding to be 'ambiguous' at background levels of ribosomal mistranslation[44,74]. Although this proposed CUG polysemy was effectively disproved, a different, spectacular case was discovered in a member of another independent yeast lineage. In *Ascoidea asiatica*, CUG is stochastically decoded by two different competing cognate tRNAs, one charged with leucine and the other with serine, resulting in nearly equal representation of both amino acids at the corresponding positions[44]. Potential codon polysemy resulting from different tRNAs competing for the same codons was once envisaged for certain nematodes but not corroborated by later proteomic work[75,76], making the situation in *A. asiatica* unique so far. A strong candidate for another case of codon polysemy nevertheless exists in some Streptomycetaceae[45,77]. These bacteria possess a member of a unique tRNA family ($tRNA^{ProX}$) bearing anticodons cognate to one of the standard codons for alanine (GCU), threonine (ACU), or asparagine (AAU), yet are charged with proline, mediated by an unusual paralog of the conventional prolyl-tRNA synthetase. Though the ability of different $tRNA^{ProX}$ to mistranslate the cognate codons as proline was demonstrated in *Escherichia coli*[77], their function in the Streptomycetaceae themselves has not been addressed directly. Hence, it is not known whether, to what extent, and under which conditions they might cause GCU, ACU, or AAU polysemy, and what the biological significance of this phenomenon might be.

Functional and evolutionary plasticity of organellar release factors

The wealth of codon reassignments found in semiautonomous organelles relies, to a large extent, on tinkering with translation termination (Figure 5). Mitochondria and plastids inherited two homologous release factors recognizing stop codons from their bacterial ancestors, referred to as mtRF1a/pRF1 (UAG/UAA) and mtRF2a/pRF2 (UAA/UGA)[78]. Loss of bacterial RF2 or its organellar counterparts was known to associate with the reassignments of UGA as a sense codon. Recently, mitochondria and plastids that have lost RF1 were reported and shown to utilize UAG (and in one case also UAA) as sense codons[79,80]. This indicates

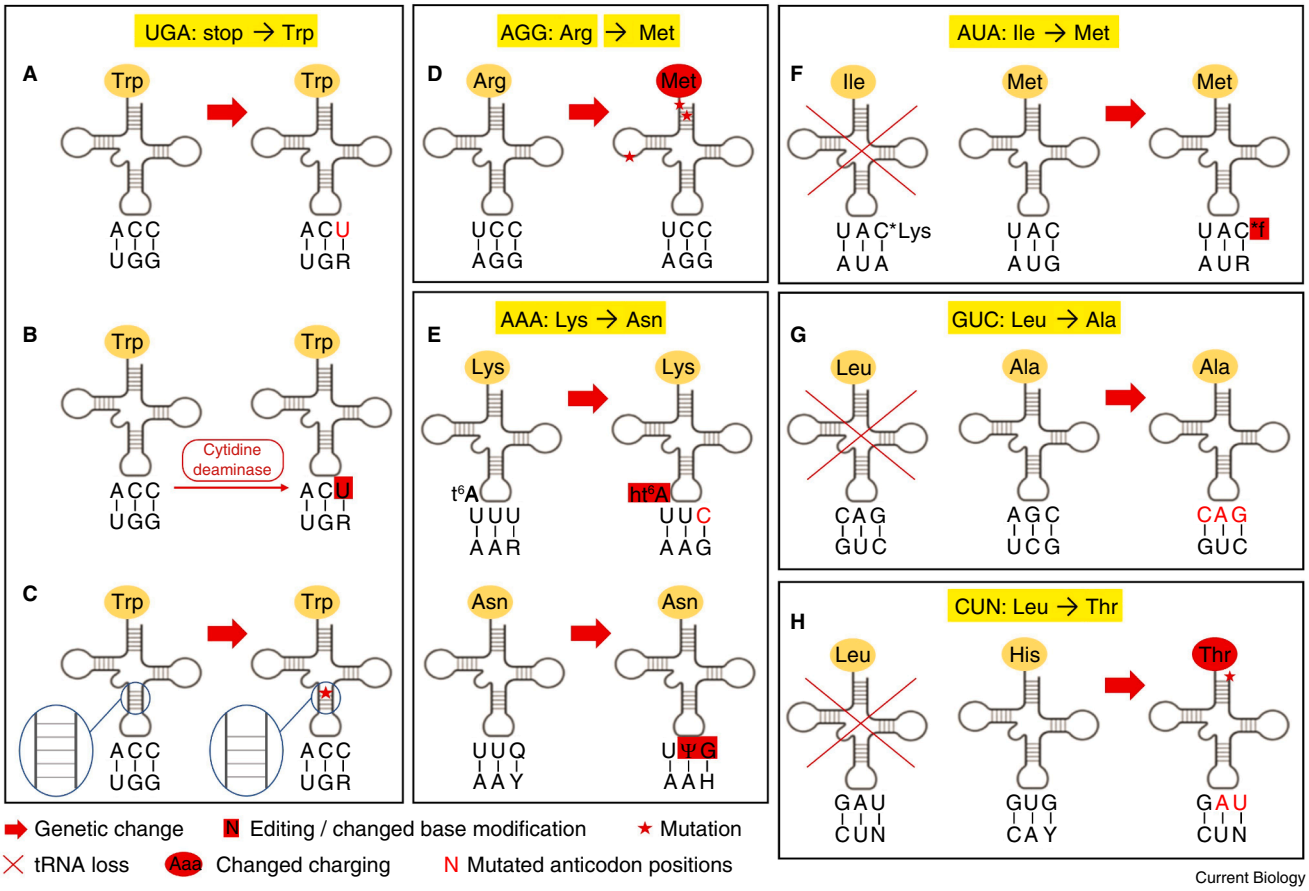


Figure 6. Types of molecular changes at the tRNA level mediating codon reassignments.
Note that the tRNAs are drawn such as to show the anticodon in the 3'-to-5' direction, whereas the codons are indicated in the 5'-to-3' direction. (A–C) Three alternative tRNA alterations that mediate UGA decoding as tryptophan in different organisms or translation systems, including mutation of the anticodon (A; for example, in human mitochondria), cytidine deamination-based editing of the anticodon (B; in trypanosomatid mitochondria), and mutation in the anticodon stem loosening the top base pair (C; for example, in cytoplasmic translation of the trypanosomatid genus *Blastocrithidia*). Note that stop-to-sense codon reassignments also require alterations in release factors (Figure 5). (D) Mutations in a tRNA may change the amino acid it is charged with, thus mediating a sense-to-sense reassignment; here exemplified by a case reported from an uncultivated lineage of the bacterial class Bacilli. (E) A sense-to-sense amino acid reassignment is mediated by changes affecting two tRNAs, one having its codon specificity restricted (top) and the other broadened (in certain invertebrate mitochondria; bottom). Changes in nucleotide modifications (both tRNAs) and mutation in the anticodon (one of the tRNAs) are involved; 'Q' denotes the modified G nucleotide queuosine. Further IUPAC base codes: H = A+C+U; Y = C+U; N = A+C+G+U (see also Figure 2). (F–H) Examples of sense-to-sense codon reassignments involving loss of the original cognate tRNAs combined with newly emerged decoding mechanisms. In the latter two cases the tRNAs with new functions represent a duplicated copy of the original tRNA indicated in the figure. (F) The Ile-to-Met AUA reassignment occurring in some mitochondria and plastids relies on an alteration of the existing $tRNA^{Met}$, for example, by formylation of its first anticodon position as displayed here. (G) CUG in the yeast order Alaninales is decoded as alanine by a tRNA evolved from a standard $tRNA^{Ala}$ by anticodon mutation. (H) The CUN codon box is decoded as threonine by an unusual tRNA evolved by transmogrification of $tRNA^{His}$ (in mitochondria of *S. cerevisiae* and relatives).

that there is no obstacle for a bacterial-type translation system to function without RF1, raising the possibility that free-living bacteria will eventually be identified with UAG reassigned as a sense codon. Indeed, this has already been achieved artificially in *E. coli*[81]. Another recent update is the discovery of the first candidate loss-of-function and gain-of-function mutations in the mitochondrial release factors mtRF1a and mtRF2a from two different protist lineages; these changes are hypothesized to restrict or broaden, respectively, codon-recognition specificity to allow for the specific stop-to-sense or sense-to-stop codon reassignments observed in these protists[79,82]. A different instance of employing non-standard termination codons (AGA and AGG) has evolved in vertebrate mitochondria. Early on, several mechanisms were proposed, but recently the vertebrate-specific mtRF1a paralog (confusingly called mtRF1) was robustly demonstrated to mediate translation termination at these codons[83–85].

The meaning of UAA and UAG may be decoupled in eukaryotes

Numerous stop-to-sense reassignments to different amino acids (Gln, Glu, or Tyr) were reported for UAG and UAA in the nuclear genomes of various protist lineages, mostly ciliates, before 2017[23,42]. In all those cases, representing at least 13 independent genetic code switches, the meaning of the two codons remained coupled, as if a functional constraint tied their fates. However, although additional cases of coupled UAG/UAA reassignment continue to be reported from protist nuclear genomes[86,87], at least seven distinct protist lineages are now

Box 3. Molecular changes behind codon reassignments.

The general framework for understanding the evolutionary steps leading to codon reassignment considers combinations of two types of evolutionary changes — losses and gains — affecting the translation machinery in variable sequences of events[63]. Below, we present examples that have been elucidated for particular codon reassignments. In this context, loss of a component of the translation system would include cases where the component has been exclusively repurposed for novel functions.

***Changes affecting release factors* (Figure 5).** Here, loss-type events include complete loss of a release factor, only possible in bacterial and organellar translation systems featuring RF1 and RF2 with partially overlapping functions, or mutations changing the codon-binding region of RFs such that their affinities for particular stop codons are diminished. These are expected to underpin stop-to-sense reassignments, and indeed outright losses of bacterial (RF2) and organellar release factors (RF1 or RF2), as well as mutations in eRF1 likely restricting its codon specificity, have been widely documented[78,209–213]. A notable example of the latter is the Ser70 mutation in eRF1, which is associated with the stop-to-Trp UGA reassignment in *Blastocrithidia*[92,94,102,212]. Candidate mutations restricting the codon-recognition specificity have been also reported in mitochondrial release factors. In mitochondria of two lineages of the green algal phylogroup Scenedesminia, UAG was independently reassigned as a sense codon (for leucine or alanine), but in this case mtRF1a could not be lost due to an earlier mtRF2a loss. Instead, a specific parallel mutation in the highly conserved codon-binding site, presumably decreasing the affinity of the protein for UAG, was noticed in mtRF1a of both lineages[82]. Another unique mutation in the codon-binding region of mtRF2a in Labyrinthulea protists likely diminishes the interaction with UAA reassigned to tyrosine[79]. Finally, a mutant form of the bacterial RF2 was engineered in *E. coli* to decrease its binding to UGA as part of a synthetic biology project[165]. Gain-type changes affecting release factors are required for sense-to-stop reassignments, which are presently known only from certain mitochondria (but see Conclusions and outlook). Here the emergence of a new release factor is well established (the vertebrate-specific mtRF1 paralog recognizing the AGR codons)[85], and candidate 'gain-of-function' mutations extending the set of codons recognized were identified in mtRF1a and mtRF2a of Scenedesminia and Labyrinthulea, respectively, to explain the use of non-standard termination codons (UCR in the former, UUA and AGR in the latter)[79,82]. However, any proposed roles for mitochondrial release factor mutations in codon-recognition changes await experimental confirmation.

***Changes affecting tRNAs* (Figure 6).** Like with release factors, tRNAs can be lost, or mutations can prevent translation of previously cognate codons leading to loss-type changes associated with sense-to-stop and sense-to-sense reassignments, both having precedents. For instance, in certain invertebrates, the standard mitochondrial $tRNA^{Lys}$ that decodes AAA/AAG as lysine has the anticodon mutated to CUU, which (together with a newly gained modification) makes it cognate to AAG only, while AAA is reassigned, decoded as asparagine by a $tRNA^{Asn}$ with a change in modification, which extends its codon-binding specificity beyond AAY[214] (Figure 6E). However, genuine tRNA losses seem to be much more common in codon reassignments. An obvious example has been mentioned in the main text as part of the Ile-to-Met reassignment in organelles: the recurrent loss of the AUA codon-specific $tRNA^{Ile}$, together with the loss of an enzyme (known as TilS) that makes it cognate to AUA instead of AUG by converting cytosine at the first anticodon position to lysidine[215]. tRNA loss is naturally also associated with sense-to-stop reassignments, exemplified by the loss of the $tRNA^{Ser}_{UGA}$ from mitochondria of Scenedesminia, where the corresponding UCR codons now signal translation termination[82]. Mirroring loss-type changes affecting tRNAs, gain-type ones may include emergence of novel tRNAs (*via* gene duplication and divergence rather than *de novo*) or acquiring the ability to pair with a particular codon by the tRNA. The stop-to-Trp UGA reassignments provide examples of different gain-type tRNA changes with this effect. These include mutation of the anticodon of the canonical $tRNA^{Trp}_{CCA}$ to UCA expanding its decoding capacity from UGG to UGA, as well as the emergence of $tRNA^{Trp}_{UCA}$ by transmogrifying a duplicated alloacceptor tRNA originally decoding a non-Trp codon and charged with a different amino acid[129]. In trypanosomatid mitochondria UGA decoding as tryptophan relies on the emergence of an organelle-targeted cytidine deaminase, which edits the anticodon of standard $tRNA^{Trp}_{CCA}$ also imported from the cytosol to UCA[216,217]. As mentioned above, another tRNA change has recently been described for *Blastocrithidia* and *Condylostoma* as part of the mechanism decoding UGA in their nuclear genes[92,94,102,218]. Here the top of the standard five nucleotide pairs forming the anticodon stem of $tRNA^{Trp}_{CCA}$ is loosened ('unpinned') due to mutation, dramatically increasing the efficiency of the tRNA as a UGA decoder. Notably, the same $tRNA^{Trp}_{CCA}$ alteration seems to have been adopted by other eukaryotes that have independently reassigned UGA to Trp in their nuclear genomes (the ciliates *Blepharisma* and *Loxodes*, and an *Amoebophrya* sp.)[104,218]. As codon–anticodon pairing is highly influenced by post-transcriptional tRNA modifications, acquiring new modifications may also take part in codon reassignments. The invertebrate mitochondrial AAA-decoding $tRNA^{Asn}$ mentioned above is one such case. Another example is provided by the aforementioned Ile-to-Met AUA reassignment in mitochondria: although decoding by a non-standard cognate $tRNA^{Met}_{UAU}$ does exist (in tunicates or bivalves)[219], a more common mechanism relies on the elongator $tRNA^{Met}_{CAU}$ that would normally decode AUG only. However, its codon specificity can be extended by modifications, which vary between taxa. In opisthokonts these include either 5-formylcytidine at the first anticodon position or a combination of an unmodified first position with N^6-threonylcarbamoyladenosine directly downstream of the anticodon[220]. How AUA decoding works in mitochondria of non-opisthokont taxa or in plastids remains to be established. In principle, a sense-to-sense codon reassignment may result from direct mutation of an extant (non-duplicated) tRNA that makes it a substrate of a different synthetase (without mutating the anticodon). This route seems to be rare, but a compelling example has been reported recently: $tRNA^{Met}_{CCU}$ evolved from $tRNA^{Arg}_{CCU}$ in an uncultivated bacterial lineage explaining its Arg-to-Met AGG reassignment[62].

(Continued on next page)

Box 3. *Continued*

Changes affecting aminoacyl-tRNA synthetases. Compared with the role of changes in release factors and tRNAs, very little is known about the role of synthetases in codon reassignments. Some synthetases do not rely on the anticodon sequence as a tRNA identity element, so in principle no alteration to their structure is needed to let them efficiently charge tRNAs base pairing with codons specifying other amino acids. This seems to be the case for SerRS or AlaRS, explaining how CUG (normally encoding leucine) can be decoded as serine or alanine with novel cognate tRNAs in yeasts displaying the different CUG reassignments[221]. In other cases, mutations in synthetases modifying their anticodon recognition are expected if they are to charge tRNAs with non-standard anticodons. This is, for instance, the case for cytosolic GlnRS in numerous eukaryotes having reassigned UAR (or UAG alone) to glutamine, but to our knowledge these have not been subject to dedicated studies. A rare exception is the mitochondrial ThrRS in a yeast clade (including *S. cerevisiae*) exhibiting a reassignment of the CUN codon box from leucine to threonine. These yeasts possess a CUN-decoding tRNA that represents a highly mutated duplicate of a canonical $tRNA^{His}$ charged by the same ThrRS as the conventional $tRNA^{Thr}_{UGU}$ decoding the standard ACN threonine codon box[222]. Experimental evidence showed that the mitochondrial ThrRS acquired mutations broadening its substrate specificity with the emergence of this novel tRNA[223]. Also notable are the discoveries in certain bacteria of novel orthogonal tRNA/synthetase pairs (consisting of mutated paralogs of standard pairs) that mediate 'deliberate mistranslation' of certain non-proline codons as proline[45,77] or UGA decoding as cysteine instead of its standard decoding as selenocysteine[119]. All the aforementioned changes affecting synthetases are of the gain-type, while we are unaware of any loss-type changes (other than those associated with 'trivial' synthetase replacements by functionally equivalent ones).

known to exhibit nuclear UAG as a sense codon (encoding Gln, Leu, or Ser), while retaining UAA exclusively as a termination codon[88–91] (Figures 2 and 4 and Table S1). This shows that eRF1 can be modified such that it retains its ability to recognize UAA while lowering its UAG affinity. Functional discrimination between UAG and UAA is also at play in the trypanosomatid genus *Blastocrithidia* or the ciliate *Peritromus kahli*. Here UAA is homonymous, retaining its translation termination function along with encoding Glu or Gln, respectively, whereas UAG is fully reassigned to that same amino acid[92–94]. Finally, UAG and UAA encoding two different amino acids (Glu and Lys, respectively), have recently been documented in an uncultivated ciliate[95].

Viral manipulation of the host genetic code

Bacteriophages utilizing UAG as a sense codon (usually for Gln) have been described, with a surprisingly high proportion of them found in gut microbiomes[96]. Strikingly, UAG reassignments have not been reported from bacteria. In addition, UGA as a sense codon has been reported not only from phages infecting bacterial hosts with the corresponding reassignment, but also from phages infecting hosts with the standard genetic code[97–99]. Hence, these phages, as well as those with UAG as a sense codon, must tweak the genetic code of their hosts to get their proteins properly translated. To this end, they frequently specify 'suppressor' tRNAs cognate to the reassigned stop codons, corresponding aminoacyl-tRNA synthetases, and/or release factors recognizing the other two (non-reassigned) stop codons[96,97,99]. The reassigned codons are typically confined to genes expressed at the late stage of phage infection, that is, only after the phage has reprogrammed its host. If any such phages use UAG or UGA as a genuine termination codon in the early-expressed genes (with stop codon recognition still being governed solely by the host's native translation system), this would imply the existence of a unique codon homonymy case, where the meaning of the codon varies between two gene categories.

Genetic codes without codons exclusively dedicated to termination

Nuclear genomes of several unrelated protist lineages departed from the standard genetic code in a truly unexpected way: they lack dedicated termination codons. Originally identified independently in certain ciliates[100,101] and *Blastocrithidia*[102], analogous code variants were subsequently found in additional ciliate lineages[93,95,103] as well as in a parasitoid dinoflagellate of the genus *Amoebophrya*[86,104]. In all cases, one, two, or all three common termination codons exhibit homonymy, specifying an amino acid when occurring in-frame and serving as a stop signal when at the end of the coding region. In at least some of these cases, such homonymous codons are depleted in the sequence preceding the bona fide stop codon[92,94,100,103], underscoring that the position along the mRNA defines the context determining the alternative interpretation of such codons. How this form of codon homonymy is implemented at the molecular level remains essentially unstudied. One of the likely explanations assumes a specific interaction between eRF1 (complexed with its regulator eukaryotic release factor 3, eRF3) and poly(A)-binding proteins. Such interactions could tether eRF1 to the end of the coding sequence near the genuine stop codon while hindering interaction with the upstream in-frame stop codons[65,100,105]. The recent establishment of *Blastocrithidia nonstop* as a genetically tractable model opens up new possibilities for experimental investigations into this question[106]. A code variant lacking dedicated termination codons has recently been reported from mitochondria of radiolarian protists, with UAG and UGA fully reassigned and UAA having a dual sense/stop meaning[107]. We note that codons with dual sense/stop meaning have been observed also in code variants that retain one or two standard termination codons, as is the case in the nuclear genome of certain ciliates[93], the plastid of a green alga[108], and a bacterial endosymbiont of insects[109]. It remains to be determined if all these cases represent codon homonymy or whether the translation apparatus, especially in the organelle and endosymbiont examples, may read the dual-meaning codons in a manner more reminiscent of codon polysemy.

The peculiar case of pyrrolysine

The discovery of co-translational insertion of pyrrolysine (Pyl) as the 22nd amino acid marks one of the most spectacular extensions of the concept of the genetic code[110,111]. However, despite more than two decades of research, important questions regarding the evolution and mechanistic implementation of Pyl

utilization remain open. Its patchy distribution in the archaeal and bacterial domains is compatible with multiple contrasting evolutionary scenarios of when Pyl utilization evolved and how its extant phylogenetic distribution was established (see Box 1 for further details). In mechanistic terms, Pyl is encoded by the UAG codon in a manner seemingly comparable with Sec encoding by UGA. However, *cis*-acting determinants analogous to SECIS that would allow the translation system to discriminate UAG instances specifying Pyl from those supposed to serve as translation stops have been postulated, but could not be corroborated[112]. Instead, it seems more likely that Pyl encoding emerges from the competition of the cognate $tRNA^{Pyl}_{CUA}$ with the respective release factor (aRF1 in archaea, RF1 in bacteria)[112,113]. This interpretation has very recently been supported by experimental evidence indicating that UAG in a model Pyl-utilizing archaeon is decoded 'ambiguously' (as Pyl or translation stop in a certain ratio of translation events)[114]. Hence, following the terminology advocated here, the encoding of Pyl (at least in the organisms where it was studied in some detail) represents a case of codon polysemy, unlike the stop/Sec UGA homonymy. The polysemous UAG decoding is very likely associated with fitness costs that seem to be mitigated in different ways. Pyl-utilizing bacteria were reported to use UAG as a common termination codon, but at least some of them express the machinery to decode it as Pyl only if specific substrates for Pyl-containing enzymes are present[115]. An analogous mechanism has also been reported from some Pyl-utilizing archaea[114]. In addition, the usage of UAG as a termination codon in these organisms tends to be low. Extreme cases of archaea using all UAG codons to encode Pyl have been reported recently; these thus exhibit a full-fledged codon reassignment[116].

Codons with a triple meaning

The early elucidation of Sec decoding of UGA, followed by the discovery of Pyl encoded in a context-dependent manner by UAG, and finally the cases of position-specific stop-codon reassignments described above provide ample evidence that a codon may signal translation termination as well as encode an amino acid in the context of the same translation system. That a homonymous codon may encode two different amino acids was first demonstrated in the ciliate *Euplotes*, in which UGA specifies cysteine (Cys) but also supports Sec insertion using the common SECIS-dependent pathway[117]. Interestingly, various bacterial lineages have switched from UGA to other codons supporting Sec insertion (still using the same general mechanism), including UAG and UAA, but also several amino acid-encoding codons[118]. Although these codons retain their standard meaning in these bacteria, they additionally become homonymous in the stop/Sec or sense/Sec configuration. Strikingly, UGA has been retained as a Sec codon also in at least some of the aforementioned eukaryotes in which this codon serves two other context-dependent roles: signalling translation termination and encoding tryptophan[86,94,100]. Thus, UGA has become triply homonymous in these organisms. A different case of a codon with triple meaning was reported from various bacteria, possessing a tRNA similar to the Sec-specific one, yet charged (or predicted to be) with cysteine, along with the standard $tRNA^{Sec}$. The ability of some of these tRNAs to decode UGA in the SECIS-dependent manner has been demonstrated in *E. coli*, suggesting that they allow the respective bacteria to produce selenoproteins alternatively with selenocysteine or cysteine in the active site, perhaps dependent on environmental cues[119,120]. This would imply UGA homonymy (stop/sense in a transcript context-dependent manner) and polysemy (Sec/Cys at the same particular positions) at the same time.

Natural genetic codes with non-triplet features

An absolute core feature of the standard genetic code is its triplet nature: coding occurs as a continuous series of non-overlapping triplet codons in a single reading frame. However, even this aspect can be tweaked by a type of recoding called 'programmed ribosomal frameshifting'[121], albeit rarely. Strikingly, ribosomal frameshifting has become systemic (that is, pangenomic) in the ciliate genus *Euplotes*[122,123], the nuclear genes of which frequently contain in-frame UAG or UAA triplets that, instead of terminating translation, signal +1 or +2 frameshifting, depending on the preceding codon. Thus, this genetic code has an inherent non-triplet feature that is apparently expanding across the genome by neutral evolution[124]. In fact, this non-canonical genetic code also lacks codons dedicated exclusively to translation termination (see above), as UGA is fully reassigned to encode Cys and UAG and UAA function as stop codons only at the very end of coding sequences[122,123]. An analogous situation has been reported from the mitochondrion of *Perkinsus* (a genus of parasitic alveolates)[125]. The synthesis of all three mitochondrially encoded proteins requires +1 or +2 ribosomal frameshifting events (one to ten per gene) at positions corresponding to specific codons that are otherwise unused in the *Perkinsus* mitogenome and thus constitute dedicated frameshifting signals.

How and why genetic codes get altered in evolution

The instances of changed codon meaning observed in nature resulted from a series of evolutionary steps, the exact order of which and the forces driving them varying among the different cases. However, specific clues are not always available to allow for a precise historical reconstruction. Still, theoretical reasoning about the possible evolutionary trajectory leading to codon reassignment and the existence of cases that seem to represent intermediate stages have led to several alternative, not necessarily mutually exclusive, scenarios that likely hold for most real-life instances.

The first explicit scenario explaining codon reassignment, 'codon capture'[126], was conceived to minimize fitness burdens for the evolving lineage. Now known as 'codon disappearance'[63], it assumes that codons first disappear from a genome (for example, because of long-term mutational bias and/or genome streamlining), followed by loss of their decoding mechanism (that is, mutation or deletion of the tRNAs and/or release factors involved). Next, a newly evolved decoding mechanism imposes new meaning for the codons upon their reappearance. Complete disappearance of codons is not unrealistic, exemplified by the complete absence of UGA from organellar genomes of diverse eukaryotes[68,127,128]. The most frequent cause of UGA disappearance is the common trend towards lowering the GC content of organellar (and other endosymbiont) genomes, resulting in UGA replacement by UAA. Interestingly, some protist mitochondria without UGA still retain the cognate mtRF2a, although mtRF1a would theoretically suffice to terminate translation here[78]. In other cases, for example, in the mitochondria of

most green algae, both UGA and mtRF2a are absent (perhaps reflecting their loss in the common ancestor of Chlorophyta)[78]. This makes these genomes primed for change: in some chlorophyte lineages UGA has reappeared as a stop codon (decoded by dually targeted pRF2 or possibly a modified mtRF1a) or became a Trp codon with the emergence of a cognate tRNA[78,82,129–131]. The latter outcome might stem from the same AT-biased mutational pressure that initially led to UGA disappearance, reintroducing the codon at the expense of the standard Trp codon UGG. Although decreasing the GC content is perhaps the most common primary cause of stop-to-Trp UGA reassignments[128], an α-proteobacterial insect endosymbiont with an extremely small genome, yet a high GC content (*Candidatus* Hodgkinia) (Figure 3), provides a notable counterexample[132]. An analogous evolutionary trajectory can result in the (for some reasons much rarer) stop-to-Trp reassignments of the UAG codon, spectacularly exemplified by the extremely AT-rich plastid genome of a non-photosynthetic parasitic plant[133].

Another codon reassignment scenario envisages an ambiguous intermediate with codons that can be decoded alternatively by both the original and newly evolved decoders[134]. A striking example has been mentioned above: the yeast *A. asiatica* with stochastic decoding of CUG by cognate $tRNA^{Leu}_{CAG}$ and $tRNA^{Ser}_{CAG}$[44]. The tRNA co-occurrence might be ancestral in the order Ascoideales, but most descendants (unlike *A. asiatica*) have since lost $tRNA^{Leu}_{CAG}$ completely or retained a mutated form with an unknown function[70,135]. An ambiguous-intermediate narrative was once invoked as an explanation for the non-monophyletic distribution of the stop-to-Gln reassignment of UAA/UAG (UAR) in the nuclear genomes of ulvophytes, assuming a reversion to the standard genetic code state in a phylogenetically nested lineage exhibiting the standard UAR meaning[136]. However, the refinement of the ulvophyte phylogeny and the discovery of a different genetic code change in a previously uncharacterized ulvophyte lineage (the stop-to-Ser reassignment of UAG in *Scotinosphaera*[89]) speak in favour of multiple completely independent genetic code changes in ulvophyte evolution (Figure 4).

Opposite the ambiguous intermediate we find the unassigned-codon scenario[63]. The original decoder (tRNA or release factor) of a codon gets lost or mutated, leaving it to be decoded suboptimally, for example, by near-cognate tRNAs, before an alternative decoding mechanism eventually arises. Though missing decoders as required by the unassigned-codon model have been reported in specific cases[128], such examples are often misleading. A case in point: the AGA and AGG codons in vertebrate mitochondria. Cognate tRNAs are absent and these codons appear only at the end of certain mitochondrial ORFs. It was initially thought that ribosomes stalled upon the appearance of these codons in the A-site would undergo a –1 frameshift, enabling termination via the standard mechanism[137], but as already mentioned, mtRF1 was eventually found as an AGA/AGG-specific release factor, terminating translation upon encountering these codons[83–85]. Changes in tRNA regions outside the anticodon with unanticipated effects on codon binding (for example, $tRNA^{Trp}_{CCA}$ in *Blastocrithidia*, certain ciliates, and an *Amoebophrya* sp.) (Box 3) or post-transcriptional tRNA alterations (by editing and/or nucleotide modifications) not readily predicted *in silico*, might be involved in decoding some other seemingly unassigned codons. Further studies into the real status of such codons, and possible alternative decoding mechanisms, are required.

In 2016, a tRNA-loss-driven scenario was proposed, in which "a tRNA or a release factor is mutated or lost, such that decoding of the cognate codon is heavily disturbed or even abolished"[42,69], which would, according to the authors, explain most known reassignments[42]. In essence, this represents an incarnation of the unassigned-codon model proposed a decade earlier. The scenario was primarily invoked to explain codon reassignments in yeasts, which were assumed to have evolved as follows: first, a single loss of $tRNA^{Leu}_{CAG}$ in a common ancestor of a broad clade of yeasts, including *S. cerevisiae* (in other words, the unassigned-codon-like situation); next, independently emerging CUG-decoding tRNAs in different descendants that, when charged with leucine, preserved the standard genetic code (as in *S. cerevisiae*), but when charged with serine or alanine gave rise to stop-to-Leu or stop-to-Ala CUG reassignments. However, detailed sequence analyses favour at least seven independent losses of the ancestral $tRNA^{Leu}_{CAG}$ gene, coupled with the emergence of alternative CUG decoders, with at least one lineage, the aforementioned order Ascoideales, following the ambiguous intermediate rather than the unassigned-codon scenario[70,135].

In light of the previous discussion, it becomes clear that formulating general conceptual frameworks for understanding deviations from the standard genetic code and reconstructing the circumstances and order of events leading up to their real-life cases are two different things. Most convincing are historical reconstructions of various mitochondrial codon reassignments[128], although usually we cannot be completely certain which of the models discussed above was followed. In particular, we may never know whether the codon was completely missing from the genome (the codon-disappearance model) or whether it was extremely rare, yet still used when the loss or gain events affecting its decoding happened. Further illustrating these difficulties, is position-specific codon decoding (codon homonymy) of the sort observed in *Blastocrithidia* or *Condylostoma* a result of an ambiguous-intermediate scenario[61]? This depends on whether the stop codons now present as in-frame sense codons were present when eRF1 still recognized them. However, if the hypothetical mechanism tethering eRF1 specifically to the end of the open reading frame (see above) evolved first, then the stop codons may have started to invade the coding sequences under circumstances more reminiscent of the codon disappearance or unassigned codon models. Notably, it was hypothesized that the ability of the translation system to 'ignore' in-frame stop codons had already evolved in the common ancestor of all ciliates[100]. It could then serve as a preadaptation allowing the system to implement the diverse stop-to-sense reassignments that have evolved in ciliates in a massively parallel manner (Figure 4). A conceptually similar but mechanistically different case of preadaptation might have facilitated the departure from the standard genetic code in the lineage leading to the genus *Blastocrithidia*. The predecessor of this and related trypanosomatids lost key components of the nonsense-mediated RNA decay pathway[138,139], possibly paving the way for the later spread of in-frame stop codons in *Blastocrithidia*[94,140].

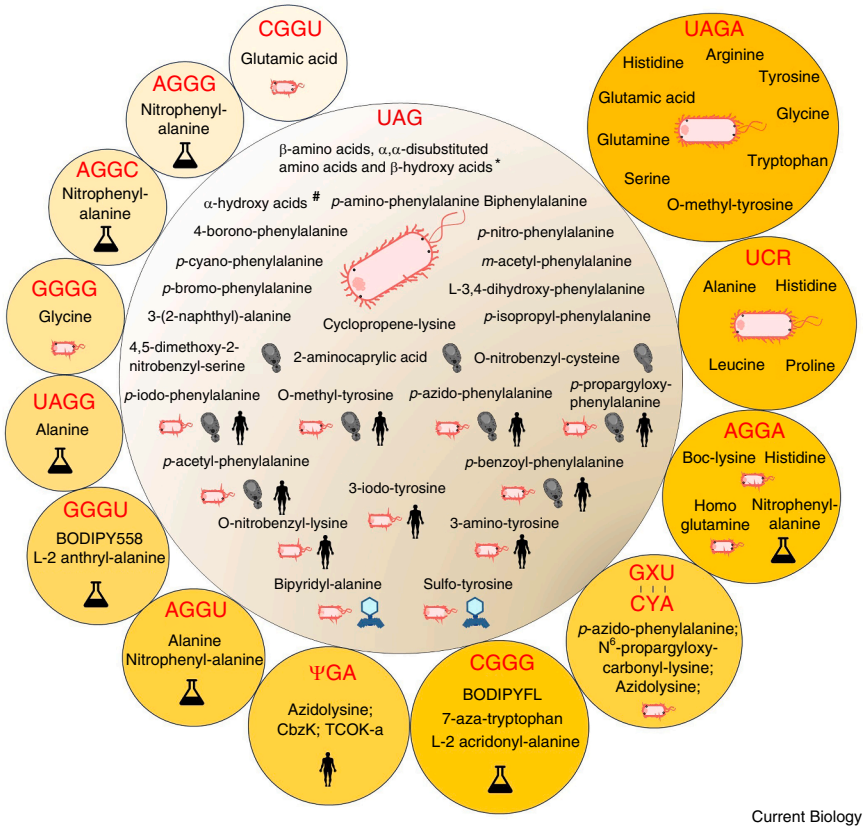
Furthermore, when speaking about codon reassignments we usually just compare these to the standard genetic code, ignoring the possibility that the extant state may result from a more complex history than simply switching from one meaning to another. Convincing cases have indeed been documented of secondary codon reassignments evolving from codons that themselves had already departed in their meaning from the standard genetic code. For example, the arginine AGA/AGG codons were most likely reassigned to serine in the mitochondrion of the common ancestor of bilaterian animals, but one or both further changed their meaning in particular bilaterian lineages to encode other amino acids or even to serve as termination codons (Figure 2)[23,128]. Another case might be the UAA/UAG (UAR) codons in the nuclear genome of peritrich ciliates encoding glutamate[141], which is usually classified as a stop-to-sense reassignment. The fact that peritrichs are phylogenetically nested among ciliate lineages all decoding UAR as glutamine suggests that evolutionarily this represents a sense-to-sense reassignment[95]. Notably, such secondary codon reassignments may in principle lead to a reversion to the standard genetic code, as seems to be the case of AUA decoded normally as isoleucine in mitochondria of certain bilaterian invertebrates despite the fact that AUA encodes methionine in most bilaterians and is likely the ancestral state for the whole group[23,128].

Finally, a salient question emerges: are codon reassignments adaptive, conferring a specific advantage to the organism, or do they reflect a non-adaptive evolutionary modality? The codon disappearance scenario can in principle be selectively neutral, with the directionality of the change dictated by a sustained mutational pressure. In contrast, the intermediate stages appearing under the ambiguous-intermediate and unassigned-codon models are expected to suffer from reduced fitness, resulting from the production of proteins with 'wrong' amino acids or inefficient decoding of particular codons during translation. Reversion to the original state (in terms of the codon meaning, but not necessarily the same mechanism achieving it) is an outcome perhaps generally favoured by natural selection and may be the most common fate, although directly observing such cases is nearly impossible. However, the selective disadvantage of intermediate stages in the ambiguous-intermediate or unassigned-codon scenarios may be relatively slim if the number of codons concerned or the number of proteins affected is low, an idea encapsulated by the proteomic constraint concept[54]. In general, the strength of selection may be at least transiently reduced depending on population genetic parameters (especially the effective population size, N_e), allowing for fixation of slightly deleterious genetic variants by genetic drift[142]. A striking demonstration of genetic drift beating out weak selection in the context of genetic code evolution was recently provided by the analysis of the spreading of the frameshift sites (manifesting a non-triplet genetic code feature, see above) in the *Euplotes* genomes[124].

Hence, codon reassignment as a result of neutral evolution is perhaps the prevailing situation, or at least the null model. Acknowledging this underscores the current general reappraisal of non-adaptationist explanations of evolutionary changes, as epitomized by the constructive neutral evolution theory[143]. Still, genetic code changes resulting from selection remain at least a theoretical possibility. The so-called 'genome streamlining hypothesis' (a term not used by the authors of the original idea themselves) supposes that codons can get reassigned due to an adaptive simplification of the translation machinery that results from selection for a smaller genome, which is assumed to operate mainly in endosymbionts and parasites[144,145]. Codon reassignments selected for as a mechanism of suppressing the effect of accidentally arising, in-frame stop codons or to better match amino acid requirements of mitochondrial proteins were also put forth[146,147]. An interesting hypothesis was proposed for the common Ile-to-Met reassignment of the AUA codon in mitochondria: specifically, that the resulting increase of the methionine content in mitochondrial proteins confers advantageous protection against oxidative stress[148]. The aforementioned recurrent loss of $tRNA^{Leu}_{CAG}$ in yeasts might be explained as a reaction to a killer-plasmid-encoded toxin specifically targeting this tRNA[70,149]. Indeed, this putative example of a genetic code change resulting from an interaction with an infectious agent might be just a specific case of a more general evolutionary response, especially to the threats posed by viruses[150]. Although viruses infecting a particular host will most often adapt to a genetic code change by co-evolving with the host lineage, organisms with a non-canonical genetic code may be protected against colonization by new viruses switching from other hosts.

Artificial genetic codes: going beyond the natural limits

Although the range of known, naturally occurring departures from the standard genetic code is stunning, the number of alternatives that are mechanistically plausible is presumed to be orders of magnitude larger. Such possible alterations include not only additional combinations of any of the proteinogenic amino acids or the translation-termination signals with particular codons, but also a whole major category of changes that would entail an expansion of the list of amino acids co-translationally inserted into nascent polypeptides beyond the known 22 natural ones (or 23, if one considers that bacteria and organelles co-translationally incorporate formylmethionine, making the textbook cliché technically incorrect). Although bioinformatic screening of prokaryotic genomes did not identify non-standard amino acids[151,152], it is worth noting that genes for unusual tRNAs and extra paralogs of various synthetases with hitherto unexplored functions have been frequently found[153,154]. Furthermore, a synthetase that charges a tRNA with a non-proteinogenic amino acid (phosphoserine) is in fact known to exist naturally, although it functions only as part of an indirect synthesis of cysteinyl-tRNA[155]. In addition, it has been known for decades that synthetases can build in structural analogues of their cognate amino acids[156]. An example is the methionine analogue azidohomoalanine, used in experimental studies with *E. coli*, although in order to achieve significant levels of its incorporation, methionine auxotrophs had to be starved of methionine first[157,158]. Using similar approaches allowed specific amino-terminal incorporation of the non-canonical methionine analogues homopropargylglycine and norleucine in almost half of a heterologously expressed protein in the yeast *S. cerevisiae*[159]. Indeed, over 90 non-canonical amino acids recognized by natural synthetases and loaded onto tRNAs have been documented[160], and most of these are translationally competent. Hence, the jury is still out on whether a mechanism for a physiologically



Current Biology

Figure 7. Artificial manipulation of the genetic code.

Several cases of incorporation of standard amino acids into proteins *via* non-natural codons, as well as of incorporation of non-canonical amino acids or even hydroxy-acids via standard or non-natural codons are shown. Efficiencies/specificities can range from polysemous codons with low artificial incorporation (compared with their natural decoding) to dedicated codons with high incorporation rates. Icons depict the systems used: human, yeast, bacterial cells, cell free extracts (represented by the flask icon); the viral icon indicates examples of phage display with non-natural amino acids. CbzK, chemically caged lysine analogue; TCOK-a, axial isomer of trans-cyclooctene-caged lysine[172]. GXU and CYA (X = NaM and Y = TPT3; chemical structure shown in Fischer *et al.*[171]) exemplify unnatural base pairs. NaM and TPT3 can interchangeably form part of the mRNA or the tRNA. NaM can pair with itself. Not all configurations work, however. *The chemical structures of the β-amino acids, α,α-disubstituted amino acids and β-hydroxy acids involved can be found in the Extended Data Figure 2 of Dunkelmann *et al.*[173]. #An early example: α-hydroxy acid p-hydroxy-L-phenyllactic acid[177]. Using cell-free extracts, many more different α-hydroxy acids can be incorporated[176]. Further information about abbreviations, references, and other details can be found in the extended list of Table S3.

relevant (adaptive) tRNA-based insertion of non-canonical amino acids into proteins has evolved in any organism.

However, what evolution has possibly failed to achieve, human creativity has managed to implement in the burgeoning synthetic biology field. Employing an ever-expanding range of approaches, synthetic biologists are capable of creating artificial genetic codes featuring unprecedented codon–amino acid couplings, highly specific genetic encoding of non-canonical amino acids, and even expanding the set of codons beyond the 64 natural ones. Admittedly, most such manipulations have so far been accomplished only in *E. coli*, but possibilities for manipulating the genetic code in mammalian cells are expanding rapidly, too. A number of reviews covering different aspects of this field have been published recently[2,3,5,6,11,161–163], so below we provide only a brief overview of artificial genetic codes, comparing and contrasting them with the natural ones where relevant.

One concern in implementing an artificial genetic code is to provide codons that would express the desired non-natural meaning. The more conventional approach employs a codon that is naturally absent from the genome or that is liberated for non-natural use by compressing the code of the model system. The latter is achieved by systematic replacement of the desired codon across the genome with a synonymous codon[164] and the deletion or modification of its cognate decoder (tRNA or release factor)[161]. The codon then may be reintroduced into a system with the new meaning, thus emulating a codon reassignment event following the codon disappearance scenario (see above). For example, such liberation of UAG and UGA stop codons was achieved in *E. coli* by replacing them with UAA combined with deletion of RF1 and introducing specific mutations into RF2, respectively[81,165]. It is noteworthy that this artificially engineered loss-of-function RF2 mutation preventing UGA decoding has no known naturally evolved parallels (Figure 5). Genome-wide compression of the genetic code, including elimination of multiple codons, has so far been reported only for *E. coli*[8,9,164,165], but there are no obvious reasons why even more massively reconfigured bacteria, and eventually other organisms, could not be viable. The prime candidate is the yeast *S. cerevisiae*[166], and work is also in progress to implement a codon-compressed plastid genome in the model green alga *Chlamydomonas reinhardtii*[167].

An alternative, more adventurous approach to genetic code expansion relies on introducing non-standard codons into the genetic system. At least three options are being developed and tested: four-base (quadruplet) codons, of which 256 versions are potentially available[168–170]; expanding the genetic alphabet by additional nucleotides that can form specific base pairs (parallel to G:C and A:T/U) and thus extra codons[11,171]; and post-transcriptional implementation of codons containing the non-standard nucleoside pseudouridine (Ψ) in an RNA-codon expansion strategy, recently implemented in mammalian cells[172]. The latter method relies on specifically designed guide snoRNAs that govern post-transcriptional conversion of uridine to Ψ in in-frame stop codons introduced at desired positions of target mRNAs, converting them to the non-standard ΨAG, ΨAA, or ΨGA codons that are then decoded by specifically designed cognate tRNAs.

Of course, not all manipulations of the genetic code bring in non-canonical amino acids, and the outcomes of such manipulations may resemble naturally evolved genetic-code variants by

having particular codons reassigned to one of the standard amino acids. For instance, the two serine codons, UCA and UCG, were artificially reassigned to encode alanine, histidine, leucine or proline in a codon-compressed *E. coli* strain provided with chimeric tRNAs charged with one of the mentioned amino acids yet having an anticodon cognate to UCA and/or UCG[9]. Notably, such specific reassignments of these codons are unknown from nature (Table S2). However, the main motivation behind tinkering with the genetic code and one of the most important promises of synthetic biology was the possibility of genetic code expansion to allow the incorporation of non-canonical amino acids. Currently, an extensive collection of over 300 such amino acids has been adopted by synthetic biologists (Figure 7 and Table S3), even including β-amino acids, which are notoriously difficult to incorporate during normal translation[163,173–175]. Further illustrating the extraordinary flexibility of ribosomal translation, α-hydroxy and α-thio acids can also be used, locally replacing peptide bonds with (thio)ester bonds in the protein backbone[176–178]. Such (amino) acids can be inserted co-translationally into growing polypeptide chains in both prokaryotes and eukaryotes and give proteins a wide range of new chemistries[11]. So far, up to four non-canonical amino acids have been efficiently incorporated in parallel by using a combination of liberated triplet codons and added quadruplet codons[173].

As a result, DNA sequence-defined, non-natural biopolymers with many novel functions may now be produced by ribosomal translation. For instance, site-specific integration of non-canonical amino acids enables: the use of photoactive amino acids for investigation of protein interactions; the use of isotopically labelled amino acids to facilitate protein structure studies; development of therapeutic vaccines aimed at breaking immunological tolerance against native proteins causing disease; production of therapeutic proteins designed to selectively target diseased tissues or tumour cells; as well as preventing protein degradation, stabilizing drugs, and building better catalysts, with new applications on the horizon[11,174,179,180]. In principle, the seemingly unlimited possibilities promised by synthetic biopolymers, primarily in medicine and bio-catalysis, could also produce fully bio-contained organisms with complete protection from viruses and organisms with the standard genetic code[11,174,179,180].

The key ingredient for genetic encoding of non-canonical amino acids is the ability to develop orthogonal molecular machineries that will ensure the desired decoding of the target codons, mainly tRNA–synthetase pairs. The number of such pairs not interfering with the use of the canonical set is increasing rapidly, providing the foundation for an ever-growing use of such compounds[181–183]. So far, up to five such mutually orthogonal pairs have been introduced in parallel into *E. coli*[181]. Generally, the best orthogonal tRNA–synthetase pairs are derived from organisms as evolutionarily distant as possible from the organism in which they are to be used, taking advantage of their extensively diverged identity elements[5,184]. Truly transformative for the field was the leveraging of the unique combination of properties of the $tRNA^{Pyl}$–synthetase pair, now used in diverse variants as the chief artificial orthogonal molecular machinery in the field (reviewed in[185]). In principle, its absence from large swaths of the tree of life secures a good chance of orthogonality in various target organisms. By default, the $tRNA^{Pyl}$ recognizes the UAG stop codon, which thus dominates among the codons used for genetic incorporation of non-canonical amino acids (Figure 7). Crucially, the synthetase tolerates anticodon changes in its substrate tRNAs, broadening the set of codons that can be exploited in this way, and has a large, promiscuous (low specificity) active site that is naturally able to charge tRNAs with multiple non-canonical amino acids. Over the years this ability has been further manipulated to extend the range of possibilities, with regard to both non-canonical amino acid incorporation as well as orthogonality[181,186,187].

Instead of only capitalizing on the evolutionarily derived divergence, physical separation might be used to obtain 'spatial orthogonality'. Examples of this can be found in what have been called semisynthetic eukaryotic cells, which allow the interference with the host translation machinery to be minimized by means of phase separation and spatial targeting. Using an RNA-targeting system it is possible to bring together a dedicated tRNA and an mRNA engineered to encode non-canonical amino acids in close proximity in a membraneless, orthogonally translating, synthetic designer 'organelle'[188]. As one of the greatest challenges in the field, certain strategies for genetic code expansion, such as the one employing quadruplet codons, require an implementation of a specially designed orthogonal ribosome operating in parallel with natural ribosomes and engineered to translate only orthogonal mRNAs[189,190]. Amazed by the sophistication of this technology, we close this brief excursion into one of the most fascinating areas of the scientific enterprise.

Conclusions and outlook

It is highly likely, especially with the large-scale genomic and proteomic studies of unicellular eukaryotes currently being performed, that new independently evolved instances of known natural code variants and novel codon reassignments will be discovered. However, we can also detect a clear tendency towards diminishing returns. It is worthwhile to point out that although 19 triplets have been found to be naturally reassigned, more than twice as many (45) have not (Figure 1). Reassignment is not random either: notably, all codons beginning with G appear untouched. The group of 19 is highly atypical: three of them are stop codons and 14 make up part of the three groups of six codon sets, encoding arginine, leucine and serine. The termination codons might be easier to change as, in principle, they come along only once per reading frame, whereas the extreme redundancy of six triplets encoding the same amino acid might accommodate occasional triplet loss in the latter group. Thus, proteomic constraints can be conquered along the lines described in this review, but most triplet meanings seem resistant to change. Underneath, however, completely different factors are often involved in such maintenance of meaning, as nature frequently exchanges random parts of complex molecular machineries[191]. This review has focused on examples that resulted in reassignment. Practically all of these alterations can be understood in a neutral framework, but selective advantages might be occasionally involved.

If random processes are able to change some genetic code meanings, then targeted development should open up codes to much more extensive manipulations. In fact, with regard to the advent of artificial codes, a revolution seems to be on the horizon. A crucial first step will be to create a 64-codon

non-degenerate code. This will allow us to encode and produce highly specific multi-functional artificial proteins with chemical capabilities and technological and therapeutic applications that are difficult to fully grasp at present. Indeed, the nature and extent of artificial codes might be hampered more by our limited imagination than by the adaptability of the translational apparatus. The future looks fruitfully flexible.

After this review was finalised and while it was being edited, several important studies in this research area became available. A study found new code variants in mitochondria of Pedinophyceae (green algae) and plastids of peDinoflagellates (having pedinophyte-derived secondary plastids), including the first sense-to-stop reassignments reported for plastids[192] (Tables S1 and S2). Another novelty: non-standard genetic codes are also employed by some of the large eukaryotic viruses of the phylum *Nucleocytoviricota*, including ones not yet reported for viruses[193]. And in the field of artificial genetic codes, Chin and co-workers generated an *E. coli* genome replacing every instance of six sense and one stop codon with synonymous codons. The resulting organism uses only 55 codons to encode the 20 canonical amino acids[194].

ACKNOWLEDGEMENTS

We thank Leoš Valášek (Institute of Microbiology, Prague) and Zdeněk Paris (Biology Centre) for discussions and Veronika Prantlová (Biology Centre) for help with the figures. We regret our inability to cover and cite many additional important studies in the field of the genetic code evolution for lack of space. This work was supported by the Czech Science Foundation grants 23-06479X (to J.L.) and 23-05764S (to M.E.), as well as the European Regional Development Fund 16_019/0000759 (to J.L. and M.E.).

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information including three tables and supplemental references can be found with this article online at https://doi.org/10.1016/j.cub.2025.09.071.

REFERENCES

1. Scerri, E. (2019). Synthetic Elements. In The Periodic Table: Its Story and Its Significance (Oxford: Oxford University Press), pp. 347–372. https://doi.org/10.1093/oso/9780190914363.003.0017.

2. Ostrov, N., Nyerges, A., Chiappino-Pepe, A., Rudolph, A., Baas-Thomas, M., and Church, G.M. (2020). Synthetic genomes with altered genetic codes. Curr. Opin. Syst. Biol. *24*, 32–40. https://doi.org/10.1016/j.coisb.2020.09.007.

3. Tang, H., Zhang, P., and Luo, X. (2022). Recent technologies for genetic code expansion and their implications on synthetic biology applications. J. Mol. Biol. *434*, 167382. https://doi.org/10.1016/j.jmb.2021.167382.

4. Mat, W.K., Xue, H., and Wong, J.T. (2010). Genetic code mutations: the breaking of a three billion year invariance. PLoS One *5*, e12206. https://doi.org/10.1371/journal.pone.0012206.

5. Chin, J.W. (2014). Expanding and reprogramming the genetic code of cells and animals. Annu. Rev. Biochem. *83*, 379–408. https://doi.org/10.1146/annurev-biochem-060713-035737.

6. Söll, D. (2015). A tRNA-guided research journey from synthetic chemistry to synthetic biology. RNA *21*, 742–744. https://doi.org/10.1261/rna.050625.115.

7. Passioura, T., and Suga, H. (2014). Reprogramming the genetic code in vitro. Trends Biochem. Sci. *39*, 400–408. https://doi.org/10.1016/j.tibs.2014.07.005.

8. Robertson, W.E., Funke, L.F.H., de la Torre, D., Fredens, J., Elliott, T.S., Spinck, M., Christova, Y., Cervettini, D., Böge, F.L., Liu, K.C., *et al*. (2021). Sense codon reassignment enables viral resistance and encoded polymer synthesis. Science *372*, 1057–1062. https://doi.org/10.1126/science.abg3029.

9. Zürcher, J.F., Robertson, W.E., Kappes, T., Petris, G., Elliott, T.S., Salmond, G.P.C., and Chin, J.W. (2022). Refactored genetic codes enable bidirectional genetic isolation. Science *378*, 516–523. https://doi.org/10.1126/science.add8943.

10. Nyerges, A., Vinke, S., Flynn, R., Owen, S.V., Rand, E.A., Budnik, B., Keen, E., Narasimhan, K., Marchand, J.A., Baas-Thomas, M., *et al*. (2023). A swapped genetic code prevents viral infections and gene transfer. Nature *615*, 720–727. https://doi.org/10.1038/s41586-023-05824-z.

11. Huang, Y., Zhang, P., Wang, H., Chen, Y., Liu, T., and Luo, X. (2025). Genetic code expansion: recent developments and emerging applications. Chem. Rev. *125*, 523–598. https://doi.org/10.1021/acs.chemrev.4c00216.

12. Brenner, S. (1957). On the impossibility of all overlapping triplet codes in information transfer from nucleic acid to proteins. Proc. Natl. Acad. Sci. USA *43*, 687–694. https://doi.org/10.1073/pnas.43.8.687.

13. Chang, J.H., and Tong, L. (2012). Mitochondrial poly(A) polymerase and polyadenylation. Biochim. Biophys. Acta *1819*, 992–997. https://doi.org/10.1016/j.bbagrm.2011.10.012.

14. Klimeš, V., Gentekaki, E., Roger, A.J., and Eliáš, M. (2014). A large number of nuclear genes in the human parasite Blastocystis require mRNA polyadenylation to create functional termination codons. Genome Biol. Evol. *6*, 1956–1961. https://doi.org/10.1093/gbe/evu146.

15. Brito Querido, J., Díaz-López, I., and Ramakrishnan, V. (2024). The molecular basis of translation initiation and its regulation in eukaryotes. Nat. Rev. Mol. Cell Biol. *25*, 168–186. https://doi.org/10.1038/s41580-023-00624-9.

16. Estrada, K., Garciarrubio, A., and Merino, E. (2024). Unraveling the plasticity of translation initiation in prokaryotes: Beyond the invariant Shine-Dalgarno sequence. PLoS One *19*, e0289914. https://doi.org/10.1371/journal.pone.0289914.

17. Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene *361*, 13–37. https://doi.org/10.1016/j.gene.2005.06.037.

18. Kearse, M.G., and Wilusz, J.E. (2017). Non-AUG translation: a new start for protein synthesis in eukaryotes. Genes Dev. *31*, 1717–1731. https://doi.org/10.1101/gad.305250.117.

19. Andreev, D.E., Loughran, G., Fedorova, A.D., Mikhaylova, M.S., Shatsky, I.N., and Baranov, P.V. (2022). Non-AUG translation initiation in mammals. Genome Biol. *23*, 111. https://doi.org/10.1186/s13059-022-02674-2.

20. Douglas, J., Cui, H., Perona, J.J., Vargas-Rodriguez, O., Tyynismaa, H., Carreño, C.A., Ling, J., Ribas de Pouplana, L., Yang, X.-L., Ibba, M., *et al*. (2024). AARS Online: A collaborative database on the structure, function, and evolution of the aminoacyl-tRNA synthetases. IUBMB Life *76*, 1091–1105. https://doi.org/10.1002/iub.2911.

21. Behrens, A., Rodschinka, G., and Nedialkova, D.D. (2021). High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. Mol. Cell *81*, 1802–1815.e7. https://doi.org/10.1016/j.molcel.2021.01.028.

22. Santos, F.B., and Del-Bem, L.E. (2022). The evolution of tRNA copy number and repertoire in cellular life. Genes (Basel) *14*, 27. https://doi.org/10.3390/genes14010027.

23. Knight, R.D., Freeland, S.J., and Landweber, L.F. (2001). Rewiring the keyboard: evolvability of the genetic code. Nat. Rev. Genet. *2*, 49–58. https://doi.org/10.1038/35047500.

24. Schultz, S.K., and Kothe, U. (2024). RNA modifying enzymes shape tRNA biogenesis and function. J. Biol. Chem. *300*, 107488. https://doi.org/10.1016/j.jbc.2024.107488.

25. Green, R., and Noller, H.F. (1997). Ribosomes and translation. Annu. Rev. Biochem. *66*, 679–716. https://doi.org/10.1146/annurev.biochem.66.1.679.

26. Mohler, K., and Ibba, M. (2017). Translational fidelity and mistranslation in the cellular response to stress. Nat. Microbiol. *2*, 17117. https://doi.org/10.1038/nmicrobiol.2017.117.

27. Ogle, J.M., and Ramakrishnan, V. (2005). Structural insights into translational fidelity. Annu. Rev. Biochem. *74*, 129–177. https://doi.org/10.1146/annurev.biochem.74.061903.155440.

28. Schwartz, M.H., and Pan, T. (2017). Function and origin of mistranslation in distinct cellular contexts. Crit. Rev. Biochem. Mol. Biol. *52*, 205–219. https://doi.org/10.1080/10409238.2016.1274284.

29. Čapková Pavlíková, Z., Miletínová, P., Roithová, A., Pospíšilová, K., Záhonová, K., Kachale, A., Becker, T., Durante, I.M., Lukeš, J., Paris, Z., *et al*. (2025). Ribosomal A-site interactions with near-cognate tRNAs drive stop codon readthrough. Nat. Struct. Mol. Biol. *32*, 662–674. https://doi.org/10.1038/s41594-024-01450-z.

30. Zhang, D., Zhu, L., Wang, F., Li, P., Wang, Y., and Gao, Y. (2023). Molecular mechanisms of eukaryotic translation fidelity and their associations with diseases. Int. J. Biol. Macromol. *242*, 124680. https://doi.org/10.1016/j.ijbiomac.2023.124680.

31. Rubio Gomez, M.A., and Ibba, M. (2020). Aminoacyl-tRNA synthetases. RNA *26*, 910–936. https://doi.org/10.1261/rna.071720.119.

32. Melnikov, S.V., Rivera, K.D., Ostapenko, D., Makarenko, A., Sanscrainte, N.D., Becnel, J.J., Solomon, M.J., Texier, C., Pappin, D.J., and Söll, D. (2018). Error-prone protein synthesis in parasites with the smallest eukaryotic genome. Proc. Natl. Acad. Sci. USA *115*, E6245–E6253. https://doi.org/10.1073/pnas.1803208115.

33. Melnikov, S.V., van den Elzen, A., Stevens, D.L., Thoreen, C.C., and Söll, D. (2018). Loss of protein synthesis quality control in host-restricted organisms. Proc. Natl. Acad. Sci. USA *115*, E11505–E11512. https://doi.org/10.1073/pnas.1815992115.

34. Hasan, F., Lant, J.T., and O'Donoghue, P. (2023). Perseverance of protein homeostasis despite mistranslation of glycine codons with alanine. Philos. Trans. R. Soc. Lond. B Biol. Sci. *378*, 20220029. https://doi.org/10.1098/rstb.2022.0029.

35. Netzer, N., Goodenbour, J.M., David, A., Dittmar, K.A., Jones, R.B., Schneider, J.R., Boone, D., Eves, E.M., Rosner, M.R., Gibbs, J.S., *et al*. (2009). Innate immune and chemically triggered oxidative stress modifies translational fidelity. Nature *462*, 522–526. https://doi.org/10.1038/nature08576.

36. Li, C., and Zhang, J. (2019). Stop-codon read-through arises largely from molecular errors and is generally nonadaptive. PLoS Genet. *15*, e1008141. https://doi.org/10.1371/journal.pgen.1008141.

37. Lynch, M.R. (2024). Evolutionary Cell Biology — The Origins of Cellular Architecture (Oxford: Oxford University Press).

38. Stadtman, T.C. (1996). Selenocysteine. Annu. Rev. Biochem. *65*, 83–100. https://doi.org/10.1146/annurev.bi.65.070196.000503.

39. Rodnina, M.V. (2023). Decoding and recoding of mRNA sequences by the ribosome. Annu. Rev. Biophys. *52*, 161–182. https://doi.org/10.1146/annurev-biophys-101922-072452.

40. Atkins, J.F., and Baranov, P.V. (2010). The distinction between recoding and codon reassignment. Genetics *185*, 1535–1536. https://doi.org/10.1534/genetics.110.119016.

41. Baranov, P.V., Atkins, J.F., and Yordanova, M.M. (2015). Augmented genetic decoding: global, local and temporal alterations of decoding processes and codon meaning. Nat. Rev. Genet. *16*, 517–529. https://doi.org/10.1038/nrg3963.

42. Kollmar, M., and Mühlhausen, S. (2017). Nuclear codon reassignments in the genomics era and mechanisms behind their evolution. BioEssays *39*, 1600221. https://doi.org/10.1002/bies.201600221.

43. Suzuki, T., Ueda, T., and Watanabe, K. (1997). The 'polysemous' codon — a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. EMBO J. *16*, 1122–1134. https://doi.org/10.1093/emboj/16.5.1122.

44. Mühlhausen, S., Schmitt, H.D., Pan, K.-T., Plessmann, U., Urlaub, H., Hurst, L.D., and Kollmar, M. (2018). Endogenous stochastic decoding of the CUG codon by competing Ser- and Leu-tRNAs in *Ascoidea asiatica*. Curr. Biol. *28*, 2046–2057. https://doi.org/10.1016/j.cub.2018.04.085.

45. Vargas-Rodriguez, O., Badran, A.H., Hoffman, K.S., Chen, M., Crnković, A., Ding, Y., Krieger, J.R., Westhof, E., Söll, D., and Melnikov, S. (2021). Bacterial translation machinery for deliberate mistranslation of the genetic code. Proc. Natl. Acad. Sci. USA *118*, e2110797118. https://doi.org/10.1073/pnas.2110797118.

46. Crick, F.H. (1968). The origin of the genetic code. J. Mol. Biol. *38*, 367–379. https://doi.org/10.1016/0022-2836(68)90392-6.

47. Buhrman, H., van der Gulik, P.T.S., Klau, G.W., Schaffner, C., Speijer, D., and Stougie, L. (2013). A realistic model under which the genetic code is optimal. J. Mol. Evol. *77*, 170–184. https://doi.org/10.1007/s00239-013-9571-2.

48. Itzkovitz, S., and Alon, U. (2007). The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res. *17*, 405–412. https://doi.org/10.1101/gr.5987307.

49. Koonin, E.V., and Novozhilov, A.S. (2017). Origin and evolution of the universal genetic code. Annu. Rev. Genet. *51*, 45–62. https://doi.org/10.1146/annurev-genet-120116-024713.

50. Rozhoňová, H., Martí-Gómez, C., McCandlish, D.M., and Payne, J.L. (2024). Robust genetic codes enhance protein evolvability. PLoS Biol. *22*, e3002594. https://doi.org/10.1371/journal.pbio.3002594.

51. Freeland, S.J., and Hurst, L.D. (1998). The genetic code is one in a million. J. Mol. Evol. *47*, 238–248. https://doi.org/10.1007/pl00006381.

52. Kisselev, L. (2002). Polypeptide release factors in prokaryotes and eukaryotes: same function, different structure. Structure *10*, 8–9. https://doi.org/10.1016/S0969-2126(01)00703-1.

53. Burroughs, A.M., and Aravind, L. (2019). The origin and evolution of release factors: implications for translation termination, ribosome rescue, and quality control pathways. Int. J. Mol. Sci. *20*, 1981. https://doi.org/10.3390/ijms20081981.

54. Massey, S.E. (2008). The proteomic constraint and its role in molecular evolution. Mol. Biol. Evol. *25*, 2557–2565. https://doi.org/10.1093/molbev/msn210.

55. Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., *et al*. (1981). Sequence and organization of the human mitochondrial genome. Nature *290*, 457–465. https://doi.org/10.1038/290457a0.

56. Barrell, B.G., Bankier, A.T., and Drouin, J. (1979). A different genetic code in human mitochondria. Nature *282*, 189–194. https://doi.org/10.1038/282189a0.

57. Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y., and Osawa, S. (1985). UGA is read as tryptophan in *Mycoplasma capricolum*. Proc. Natl. Acad. Sci. USA *82*, 2306–2309. https://doi.org/10.1073/pnas.82.8.2306.

58. Preer, Jr., J.R., Preer, L.B., Rudman, B.M., and Barnett, A.J. (1985). Deviation from the universal code shown by the gene for surface protein 51A in *Paramecium*. Nature *314*, 188–190. https://doi.org/10.1038/314188a0.

59. Renaudin, J., Pascarel, M.C., and Bové, J.M. (1987). Spiroplasma virus 4: nucleotide sequence of the viral DNA, regulatory signals, and proposed genome organization. J. Bacteriol. *169*, 4950–4961. https://doi.org/10.1128/jb.169.11.4950-4961.1987.

60. Lang-Unnasch, N., and Aiello, D.P. (1999). Sequence evidence for an altered genetic code in the *Neospora caninum* plastid. Int. J. Parasitol. *29*, 1557–1562. https://doi.org/10.1016/s0020-7519(99)00119-8.

61. Keeling, P.J. (2016). Genomics: evolution of the genetic code. Curr. Biol. *26*, R851–R853. https://doi.org/10.1016/j.cub.2016.08.005.

62. Shulgina, Y., and Eddy, S.R. (2021). A computational screen for alternative genetic codes in over 250,000 genomes. eLife *10*, e71402. https://doi.org/10.7554/eLife.71402.

63. Sengupta, S., and Higgs, P.G. (2005). A unified model of codon reassignment in alternative genetic codes. Genetics *170*, 831–840. https://doi.org/10.1534/genetics.104.037887.

64. Ling, J., O'Donoghue, P., and Söll, D. (2015). Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology. Nat. Rev. Microbiol. *13*, 707–721. https://doi.org/10.1038/nrmicro3568.

65. Salman, A., Biziaev, N., Shuvalova, E., and Alkalaeva, E. (2024). mRNA context and translation factors determine decoding in alternative nuclear genetic codes. BioEssays *46*, e2400058. https://doi.org/10.1002/bies.202400058.

66. Santos, M.A., Gomes, A.C., Santos, M.C., Carreto, L.C., and Moura, G.R. (2011). The genetic code of the fungal CTG clade. C. R. Biol. *334*, 607–611. https://doi.org/10.1016/j.crvi.2011.05.008.

67. Matsumoto, T., Ishikawa, S.A., Hashimoto, T., and Inagaki, Y. (2011). A deviant genetic code in the green alga-derived plastid in the dinoflagellate *Lepidodinium chlorophorum*. Mol. Phylogenet. Evol. *60*, 68–72. https://doi.org/10.1016/j.ympev.2011.04.010.

68. Turmel, M., Lopes Dos Santos, A., Otis, C., Sergerie, R., and Lemieux, C. (2019). Tracing the evolution of the plastome and mitogenome in the Chloropicophyceae uncovered convergent tRNA gene losses and a variant plastid genetic code. Genome Biol. Evol. *11*, 1275–1292. https://doi.org/10.1093/gbe/evz074.

69. Mühlhausen, S., Findeisen, P., Plessmann, U., Urlaub, H., and Kollmar, M. (2016). A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. Genome Res. *26*, 945–955. https://doi.org/10.1101/gr.200931.115.

70. Krassowski, T., Coughlan, A.Y., Shen, X.X., Zhou, X., Kominek, J., Opulente, D.A., Riley, R., Grigoriev, I.V., Maheshwari, N., Shields, D.C., *et al*. (2018). Evolutionary instability of CUG-Leu in the genetic code of budding yeasts. Nat. Commun. *9*, 1887. https://doi.org/10.1038/s41467-018-04374-7.

71. Santos, M.A., Ueda, T., Watanabe, K., and Tuite, M.F. (1997). The non-standard genetic code of *Candida* spp.: an evolving genetic code or a novel mechanism for adaptation? Mol. Microbiol. *26*, 423–431. https://doi.org/10.1046/j.1365-2958.1997.5891961.x.

72. Miranda, I., Silva-Dias, A., Rocha, R., Teixeira-Santos, R., Coelho, C., Gonçalves, T., Santos, M.A., Pina-Vaz, C., Solis, N.V., Filler, S.G., and Rodrigues, A.G. (2013). *Candida albicans* CUG mistranslation is a mechanism to create cell surface variation. mBio *4*, e00285-13. https://doi.org/10.1128/mBio.00285-13.

73. Gomes, A.C., Miranda, I., Silva, R.M., Moura, G.R., Thomas, B., Akoulitchev, A., and Santos, M.A.S. (2007). A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. Genome Biol. *8*, R206. https://doi.org/10.1186/gb-2007-8-10-r206.

74. Mühlhausen, S., Schmitt, H.D., Plessmann, U., Mienkus, P., Sternisek, P., Perl, T., Weig, M., Urlaub, H., Bader, O., and Kollmar, M. (2021). Proteogenomics analysis of CUG codon translation in the human pathogen *Candida albicans*. BMC Biol. *19*, 258. https://doi.org/10.1186/s12915-021-01197-9.

75. Hamashima, K., Fujishima, K., Masuda, T., Sugahara, J., Tomita, M., and Kanai, A. (2011). Nematode-specific tRNAs that decode an alternative genetic code for leucine. Nucleic Acids Res. *40*, 3653–3662. https://doi.org/10.1093/nar/gkr1226.

76. Hamashima, K., Mori, M., Andachi, Y., Tomita, M., Kohara, Y., and Kanai, A. (2015). Analysis of genetic code ambiguity arising from nematode-specific misacylated tRNAs. PLoS One *10*, e0116981. https://doi.org/10.1371/journal.pone.0116981.

77. Schuntermann, D.B., Fischer, J.T., Bile, J., Gaier, S.A., Shelley, B.A., Awawdeh, A., Jahn, M., Hoffman, K.S., Westhof, E., Söll, D., *et al*. (2023). Mistranslation of the genetic code by a new family of bacterial transfer RNAs. J. Biol. Chem. *299*, 104852. https://doi.org/10.1016/j.jbc.2023.104852.

78. Duarte, I., Nabuurs, S.B., Magno, R., and Huynen, M. (2012). Evolution and diversification of the organellar release factor family. Mol. Biol. Evol. *29*, 3497–3512. https://doi.org/10.1093/molbev/mss157.

79. Žihala, D., Salamonová, J., and Eliáš, M. (2020). Evolution of the genetic code in the mitochondria of Labyrinthulea (Stramenopiles). Mol. Phylogenet. Evol. *152*, 106908. https://doi.org/10.1016/j.ympev.2020.106908.

80. Ceriotti, L.F., Roulet, M.E., and Sanchez-Puerta, M.V. (2021). Plastomes in the holoparasitic family Balanophoraceae: Extremely high AT content, severe gene content reduction, and two independent genetic code changes. Mol. Phylogenet. Evol. *162*, 107208. https://doi.org/10.1016/j.ympev.2021.107208.

81. Lajoie, M.J., Rovner, A.J., Goodman, D.B., Aerni, H.-R., Haimovich, A.D., Kuznetsov, G., Mercer, J.A., Wang, H.H., Carr, P.A., Mosberg, J.A., *et al*. (2013). Genomically recoded organisms expand biological functions. Science *342*, 357–360. https://doi.org/10.1126/science.1241459.

82. Žihala, D., and Eliáš, M. (2019). Evolution and unprecedented variants of the mitochondrial genetic code in a lineage of green algae. Genome Biol. Evol. *11*, 2992–3007. https://doi.org/10.1093/gbe/evz210.

83. Krüger, A., Remes, C., Shiriaev, D.I., Liu, Y., Spåhr, H., Wibom, R., Atanassov, I., Nguyen, M.D., Cooperman, B.S., and Rorbach, J. (2023). Human mitochondria require mtRF1 for translation termination at non-canonical stop codons. Nat. Commun. *14*, 30. https://doi.org/10.1038/s41467-022-35684-6.

84. Saurer, M., Leibundgut, M., Nadimpalli, H.P., Scaiola, A., Schönhut, T., Lee, R.G., Siira, S.J., Rackham, O., Dreos, R., Lenarčič, T., *et al*. (2023). Molecular basis of translation termination at noncanonical stop codons in human mitochondria. Science *380*, 531–536. https://doi.org/10.1126/science.adf9890.

85. Krüger, A., Kovalchuk, D., Shiriaev, D., and Rorbach, J. (2024). Decoding the enigma: translation termination in human mitochondria. Hum. Mol. Genet. *33*, R42–R46. https://doi.org/10.1093/hmg/ddae032.

86. Bachvaroff, T.R. (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*, e0212912. https://doi.org/10.1371/journal.pone.0212912.

87. Rotterová, J., Pánek, T., Salomaki, E.D., Kotyk, M., Táborský, P., Kolísko, M., and Čepička, I. (2024). Single cell transcriptomics reveals UAR codon reassignment in *Palmarella salina* (Metopida, Armophorea) and confirms Armophorida belongs to APM clade. Mol. Phylogenet. Evol. *191*, 107991. https://doi.org/10.1016/j.ympev.2023.107991.

88. Pánek, T., Žihala, D., Sokol, M., Derelle, R., Klimeš, V., Hradilová, M., Zadrobílková, E., Susko, E., Roger, A.J., Čepička, I., and Eliáš, M. (2017). Nuclear genetic codes with a different meaning of the UAG and the UAA codon. BMC Biol. *15*, 8. https://doi.org/10.1186/s12915-017-0353-y.

89. Li, X., Hou, Z., Xu, C., Shi, X., Yang, L., Lewis, L.A., and Zhong, B. (2021). Large phylogenomic data sets reveal deep relationships and trait evolution in chlorophyte green algae. Genome Biol. Evol. *13*, evab101. https://doi.org/10.1093/gbe/evab101.

90. McGowan, J., Richards, T.A., Hall, N., and Swarbreck, D. (2024). Multiple independent genetic code reassignments of the UAG stop codon in phyllopharyngean ciliates. PLoS Genet. *20*, e1011512. https://doi.org/10.1371/pgen.1011512.

91. Pánek, T., Tice, A.K., Corre, P., Hrubá, P., Žihala, D., Kamikawa, R., Yazaki, E., Shiratori, T., Kume, K., Hashimoto, T., *et al*. (2025). An expanded phylogenomic analysis of Heterolobosea reveals the deep relationships, non-canonical genetic codes, and cryptic flagellate stages in the group. Mol. Phylogenet. Evol. *204*, 108289. https://doi.org/10.1016/j.ympev.2025.108289.

92. Kachale, A., Pavlíková, Z., Nenarokova, A., Roithová, A., Durante, I.M., Miletínová, P., Záhonová, K., Nenarokov, S., Votýpka, J., Horáková, E., *et al*. (2023). Short tRNA anticodon stem and mutant eRF1 allow stop codon reassignment. Nature *613*, 751–758. https://doi.org/10.1038/s41586-022-05584-2.

93. Chen, W., Geng, Y., Zhang, B., Yan, Y., Zhao, F., and Miao, M. (2023). Stop or not: genome-wide profiling of reassigned stop codons in ciliates. Mol. Biol. Evol. *40*, msad064. https://doi.org/10.1093/molbev/msad064.

94. Záhonová, K., Füssy, Z., Albanaz, A., Butenko, A., Kachale, A., Kraeva, N., Galan, A., Zakharova, A., Stojanova, B., Votýpka, J., *et al*. (2025). Comparative genomic analysis of trypanosomatid protists illuminates an extensive change in the nuclear genetic code. mBio *16*, e00885-25. https://doi.org/10.1128/mbio.00885-25.

95. McGowan, J., Kilias, E.S., Alacid, E., Lipscombe, J., Jenkins, B.H., Gharbi, K., Kaithakottil, G.G., Macaulay, I.C., McTaggart, S., Warring, S.D., *et al*. (2023). Identification of a non-canonical ciliate nuclear genetic code where UAA and UAG code for different amino acids. PLoS Genet. *19*, e1010913. https://doi.org/10.1371/journal.pgen.1010913.

96. Ivanova, N.N., Schwientek, P., Tripp, H.J., Rinke, C., Pati, A., Huntemann, M., Visel, A., Woyke, T., Kyrpides, N.C., and Rubin, E.M. (2014). Stop codon reassignments in the wild. Science *344*, 909–913. https://doi.org/10.1126/science.1250691.

97. Yutin, N., Benler, S., Shmakov, S.A., Wolf, Y.I., Tolstoy, I., Rayko, M., Antipov, D., Pevzner, P.A., and Koonin, E.V. (2021). Analysis of metagenome-assembled viral genomes from the human gut reveals diverse putative CrAss-like phages with unique genomic features. Nat. Commun. *12*, 1044. https://doi.org/10.1038/s41467-021-21350-w.

98. Pfennig, A., Lomsadze, A., and Borodovsky, M. (2023). MgCod: gene prediction in phage genomes with multiple genetic codes. J. Mol. Biol. *435*, 168159. https://doi.org/10.1016/j.jmb.2023.168159.

99. Borges, A.L., Lou, Y.C., Sachdeva, R., Al-Shayeb, B., Penev, P.I., Jaffe, A.L., Lei, S., Santini, J.M., and Banfield, J.F. (2022). Widespread stop-codon recoding in bacteriophages may regulate translation of lytic genes. Nat. Microbiol. *7*, 918–927. https://doi.org/10.1038/s41564-022-01128-6.

100. Swart, E.C., Serra, V., Petroni, G., and Nowacki, M. (2016). Genetic codes with no dedicated stop codon: context-dependent translation termination. Cell *166*, 691–702. https://doi.org/10.1016/j.cell.2016.06.020.

101. Heaphy, S.M., Mariotti, M., Gladyshev, V.N., Atkins, J.F., and Baranov, P. V. (2016). Novel ciliate genetic code variants including the reassignment of all three stop codons to sense codons in *Condylostoma magnum*. Mol. Biol. Evol. *33*, 2885–2889. https://doi.org/10.1093/molbev/msw166.

102. Záhonová, K., Kostygov, A.Y., Ševčíková, T., Yurchenko, V., and Eliáš, M. (2016). An unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. Curr. Biol. *26*, 2364–2369. https://doi.org/10.1016/j.cub.2016.06.064.

103. Seah, B.K.B., Singh, A., and Swart, E.C. (2022). Karyorelict ciliates use an ambiguous genetic code with context-dependent stop/sense codons. Peer Community J. *2*, e42. https://doi.org/10.24072/pcjournal.141.

104. DeMontigny, W., and Bachvaroff, T. (2025). The nuclear and mitochondrial genomes of *amoebophrya* sp. ex *Karlodinium veneficum*. G3 (Bethesda) *15*, jkaf030. https://doi.org/10.1093/g3journal/jkaf030.

105. Alkalaeva, E., and Mikhailova, T. (2017). Reassigning stop codons via translation termination: How a few eukaryotes broke the dogma. BioEssays *39*, 1600213. https://doi.org/10.1002/bies.201600213.

106. Galan, A., Kraeva, N., Záhonová, K., Butenko, A., Kostygov, A.Y., Paris, Z., Pergner, J., Bianchi, C., Fakih, F., Saura, A., *et al*. (2025). Converting *Blastocrithidia nonstop*, a trypanosomatid with non-canonical genetic code, into a genetically-tractable model. Mol. Microbiol. *123*, 586–592. https://doi.org/10.1111/mmi.15365.

107. Macher, J.-N., Coots, N.L., Poh, Y.-P., Girard, E.B., Langerak, A., Muñoz-Gómez, S.A., Sinha, S.D., Jirsová, D., Vos, R., Wissels, R., *et al*. (2023). Single-cell genomics reveals the divergent mitochondrial genomes of Retaria (Foraminifera and Radiolaria). mBio *14*. e00302–23. https://doi.org/10.1128/mbio.00302-23.

108. Del Cortona, A., Leliaert, F., Bogaert, K.A., Turmel, M., Boedeker, C., Janouškovec, J., Lopez-Bautista, J.M., Verbruggen, H., Vandepoele, K., and De Clerck, O. (2017). The plastid genome in Cladophorales green algae is encoded by hairpin chromosomes. Curr. Biol. *27*, 3771–3782. https://doi.org/10.1016/j.cub.2017.11.004.

109. Manzano-Marín, A., Kvist, S., and Oceguera-Figueroa, A. (2023). Evolution of an alternative genetic code in the *Providencia* symbiont of the hematophagous leech *Haementeria acuecueyetzin*. Genome Biol. Evol. *15*, evad164. https://doi.org/10.1093/gbe/evad164.

110. Srinivasan, G., James, C.M., and Krzycki, J.A. (2002). Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science *296*, 1459–1462. https://doi.org/10.1126/science.1069588.

111. Hao, B., Gong, W., Ferguson, T.K., James, C.M., Krzycki, J.A., and Chan, M.K. (2002). A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science *296*, 1462–1466. https://doi.org/10.1126/science.1069556.

112. Brugère, J.F., Atkins, J.F., O'Toole, P.W., and Borrel, G. (2018). Pyrrolysine in archaea: a 22nd amino acid encoded through a genetic code expansion. Emerg. Top. Life Sci. *2*, 607–618. https://doi.org/10.1042/etls20180094.

113. Alkalaeva, E., Eliseev, B., Ambrogelly, A., Vlasov, P., Kondrashov, F.A., Gundllapalli, S., Frolova, L., Söll, D., and Kisselev, L. (2009). Translation termination in pyrrolysine-utilizing archaea. FEBS Lett. *583*, 3455–3460. https://doi.org/10.1016/j.febslet.2009.09.044.

114. Shalvarjian, K.E., Chadwick, G.L., Pérez, P.I., Woods, P.H., Orphan, V.J., and Nayak, D.D. (2025). The ambiguous genetic code of methanogenic archaea that grow on methylamines. Preprint at bioRxiv, https://doi.org/10.1101/2025.06.11.659114.

115. Prat, L., Heinemann, I.U., Aerni, H.R., Rinehart, J., O'Donoghue, P., and Söll, D. (2012). Carbon source-dependent expansion of the genetic code in bacteria. Proc. Natl. Acad. Sci. USA *109*, 21070–21075. https://doi.org/10.1073/pnas.1218613110.

116. Kivenson, V., Peters, S.L., Borrel, G., Kivenson, A., Roe, L.T., Hamlish, N. X., Fadhlaoui, K., Schepartz, A., Gribaldo, S., Hettich, R.L., and Banfield, J.F. (2024). Sporadic distribution of a new archaeal genetic code with all TAG codons as pyrrolysine. Preprint at bioRxiv, https://doi.org/10.1101/2024.09.30.615893.

117. Turanov, A.A., Lobanov, A.V., Fomenko, D.E., Morrison, H.G., Sogin, M.L., Klobutcher, L.A., Hatfield, D.L., and Gladyshev, V.N. (2009). Genetic code supports targeted insertion of two amino acids by one codon. Science *323*, 259–261. https://doi.org/10.1126/science.1164748.

118. Mukai, T., Englert, M., Tripp, H.J., Miller, C., Ivanova, N.N., Rubin, E.M., Kyrpides, N.C., and Söll, D. (2016). Facile recoding of selenocysteine in nature. Angew. Chem. *55*, 5337–5341. https://doi.org/10.1002/anie.201511657.

119. Mukai, T., Vargas-Rodriguez, O., Englert, M., Tripp, H.J., Ivanova, N.N., Rubin, E.M., Kyrpides, N.C., and Söll, D. (2016). Transfer RNAs with novel cloverleaf structures. Nucleic Acids Res. *45*, 2776–2785. https://doi.org/10.1093/nar/gkw898.

120. Vargas-Rodriguez, O., Englert, M., Merkuryev, A., Mukai, T., and Söll, D. (2018). Recoding of the selenocysteine UGA codon by cysteine in the presence of a non-canonical $tRNA^{Cys}$ and elongation factor SelB. RNA Biol. *15*, 471–479. https://doi.org/10.1080/15476286.2018.1474074.

121. Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E., and Baranov, P.V. (2016). Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. Nucleic Acids Res. *44*, 7007–7078. https://doi.org/10.1093/nar/gkw530.

122. Klobutcher, L.A., and Farabaugh, P.J. (2002). Shifty ciliates: frequent programmed translational frameshifting in euplotids. Cell *111*, 763–766. https://doi.org/10.1016/S0092-8674(02)01138-8.

123. Lobanov, A.V., Heaphy, S.M., Turanov, A.A., Gerashchenko, M.V., Pucciarelli, S., Devaraj, R.R., Xie, F., Petyuk, V.A., Smith, R.D., Klobutcher, L.A., *et al*. (2017). Position-dependent termination and widespread obligatory frameshifting in *Euplotes* translation. Nat. Struct. Mol. Biol. *24*, 61–68. https://doi.org/10.1038/nsmb.3330.

124. Gaydukova, S.A., Moldovan, M.A., Vallesi, A., Heaphy, S.M., Atkins, J.F., Gelfand, M.S., and Baranov, P.V. (2023). Nontriplet feature of genetic code in *Euplotes* ciliates is a result of neutral evolution. Proc. Natl. Acad. Sci. USA *120*, e2221683120. https://doi.org/10.1073/pnas.2221683120.

125. Gornik, S.G., Flores, V., Reinhardt, F., Erber, L., Salas-Leiva, D.E., Douvropoulou, O., Lassadi, I., Einarsson, E., Mörl, M., Git, A., *et al*. (2022). Mitochondrial genomes in Perkinsus decode conserved frameshifts in all genes. Mol. Biol. Evol. *39*, msac191. https://doi.org/10.1093/molbev/msac191.

126. Osawa, S., and Jukes, T.H. (1989). Codon reassignment (codon capture) in evolution. J. Mol. Evol. *28*, 271–278. https://doi.org/10.1007/bf02103422.

127. Ševčíková, T., Klimeš, V., Zbránková, V., Strnad, H., Hroudová, M., Vlček, Č., and Eliáš, M. (2016). A comparative analysis of mitochondrial genomes in eustigmatophyte algae. Genome Biol. Evol. *8*, 705–722. https://doi.org/10.1093/gbe/evw027.

128. Sengupta, S., Yang, X., and Higgs, P.G. (2007). The mechanisms of codon reassignments in mitochondrial genetic codes. J. Mol. Evol. *64*, 662–688. https://doi.org/10.1007/s00239-006-0284-7.

129. Noutahi, E., Calderon, V., Blanchette, M., El-Mabrouk, N., and Lang, B.F. (2019). Rapid genetic code evolution in green algal mitochondrial genomes. Mol. Biol. Evol. *36*, 766–783. https://doi.org/10.1093/molbev/msz016.

130. Turmel, M., Vincent, A.T., Otis, C., and Lemieux, C. (2020). The complete plastome of the coccoid green alga *Jenufa minuta* (Chlorophyceae, incertae sedis) unveils a noncanonical genetic code and a previously unrecognized trans-spliced group II intron in the rpl32 gene. Mitochondrial DNA B *5*, 1728–1730. https://doi.org/10.1080/23802359.2020.1749165.

131. Hulatt, C.J., Suzuki, H., Détain, A., Wijffels, R.H., Leya, T., and Posewitz, M.C. (2024). The genome of the Arctic snow alga *Limnomonas spitsbergensis* (Chlamydomonadales). G3 (Bethesda) *14*, jkae086. https://doi.org/10.1093/g3journal/jkae086.

132. McCutcheon, J.P., McDonald, B.R., and Moran, N.A. (2009). Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont. PLoS Genet. *5*, e1000565. https://doi.org/10.1371/journal.pgen.1000565.

133. Su, H.J., Barkman, T.J., Hao, W., Jones, S.S., Naumann, J., Skippington, E., Wafula, E.K., Hu, J.M., Palmer, J.D., and dePamphilis, C.W. (2019). Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant *Balanophora*. Proc. Natl. Acad. Sci. USA *116*, 934–943. https://doi.org/10.1073/pnas.1816822116.

134. Schultz, D.W., and Yarus, M. (1994). Transfer RNA mutation and the malleability of the genetic code. J. Mol. Biol. *235*, 1377–1380. https://doi.org/10.1006/jmbi.1994.1094.

135. Cinnéide, E.Ó., Scaife, C., Dillon, E.T., and Wolfe, K.H. (2024). Evolution of the genetic code in the Ascoideales (CUG-Ser2) yeast clade: The ancestral tRNA-Leu(CAG) gene is retained in most *Saccharomycopsis* species but is nonessential and not used for translation. Genome Biol. Evol. *16*, evae166. https://doi.org/10.1093/gbe/evae166.

136. Cocquyt, E., Gile, G.H., Leliaert, F., Verbruggen, H., Keeling, P.J., and De Clerck, O. (2010). Complex phylogenetic distribution of a non-canonical genetic code in green algae. BMC Evol. Biol. *10*, 327. https://doi.org/10.1186/1471-2148-10-327.

137. Temperley, R., Richter, R., Dennerlein, S., Lightowlers, R.N., and Chrzanowska-Lightowlers, Z.M. (2010). Hungry codons promote frameshifting in human mitochondrial ribosomes. Science *327*, 301. https://doi.org/10.1126/science.1180674.

138. Delhi, P., Queiroz, R., Inchaustegui, D., Carrington, M., and Clayton, C. (2011). Is there a classical nonsense-mediated decay pathway in trypanosomes? PLoS One *6*, e25112. https://doi.org/10.1371/journal.pone.0025112.

139. Nenarokova, A., Záhonová, K., Krasilnikova, M., Gahura, O., McCulloch, R., Zíková, A., Yurchenko, V., and Lukeš, J. (2019). Causes and effects of loss of classical nonhomologous end joining pathway in parasitic eukaryotes. mBio *10*, e01541-19. https://doi.org/10.1128/mBio.01541-19.

140. Lukeš, J., Čapková Pavlíková, Z., Yurchenko, V., Paris, Z., and Valášek, L.S. (2025). *Blastocrithidia* — a genetic alien from the planet Earth. Cold Spring Harb. Perspect. Biol. a041868. https://doi.org/10.1101/cshperspect.a041868.

141. Sánchez-Silva, R., Villalobo, E., Morin, L., and Torres, A. (2003). A new noncanonical nuclear genetic code: translation of UAA into glutamate. Curr. Biol. *13*, 442–447. https://doi.org/10.1016/s0960-9822(03)00126-x.

142. Lynch, M., Bobay, L.M., Catania, F., Gout, J.F., and Rho, M. (2011). The repatterning of eukaryotic genomes by random genetic drift. Annu. Rev. Genom. Hum. Genet. *12*, 347–366. https://doi.org/10.1146/annurev-genom-082410-101412.

143. Lukeš, J., Archibald, J.M., Keeling, P.J., Doolittle, W.F., and Gray, M.W. (2011). How a neutral evolutionary ratchet can build cellular complexity. IUBMB Life *63*, 528–537. https://doi.org/10.1002/iub.489.

144. Andersson, S.G., and Kurland, C.G. (1995). Genomic evolution drives the evolution of the translation system. Biochem. Cell Biol. *73*, 775–787. https://doi.org/10.1139/o95-086.

145. Andersson, S.G.E., and Kurland, C.G. (1998). Reductive evolution of resident genomes. Trends Microbiol. *6*, 263–268. https://doi.org/10.1016/S0966-842X(98)01312-2.

146. Johnson, L.J. (2010). Pseudogene rescue: an adaptive mechanism of codon reassignment. J. Evol. Biol. *23*, 1623–1630. https://doi.org/10.1111/j.1420-9101.2010.02027.x.

147. Swire, J., Judson, O.P., and Burt, A. (2005). Mitochondrial genetic codes evolve to match amino acid requirements of proteins. J. Mol. Evol. *60*, 128–139. https://doi.org/10.1007/s00239-004-0077-9.

148. Bender, A., Hajieva, P., and Moosmann, B. (2008). Adaptive antioxidant methionine accumulation in respiratory chain complexes explains the use of a deviant genetic code in mitochondria. Proc. Natl. Acad. Sci. USA *105*, 16496–16501. https://doi.org/10.1073/pnas.0802779105.

149. Heneghan, P.G., Salzberg, L.I., Ó Cinnéide, E., Dewald, J.A., Weinberg, C.E., and Wolfe, K.H. (2025). Ancient origin and high diversity of zymocin-like killer toxins in the budding yeast subphylum. Proc. Natl. Acad. Sci. USA *122*, e2419860122. https://doi.org/10.1073/pnas.2419860122.

150. Shackelton, L.A., and Holmes, E.C. (2008). The role of alternative genetic codes in viral evolution and emergence. J. Theor. Biol. *254*, 128–134. https://doi.org/10.1016/j.jtbi.2008.05.024.

151. Lobanov, A.V., Kryukov, G.V., Hatfield, D.L., and Gladyshev, V.N. (2006). Is there a twenty third amino acid in the genetic code? Trends Genet. *22*, 357–360. https://doi.org/10.1016/j.tig.2006.05.002.

152. Fujita, M., Mihara, H., Goto, S., Esaki, N., and Kanehisa, M. (2007). Mining prokaryotic genomes for unknown amino acids: a stop-codon-based approach. BMC Bioinformatics *8*, 225. https://doi.org/10.1186/1471-2105-8-225.

153. Chaliotis, A., Vlastaridis, P., Mossialos, D., Ibba, M., Becker, H.D., Stathopoulos, C., and Amoutzias, G.D. (2016). The complex evolutionary history of aminoacyl-tRNA synthetases. Nucleic Acids Res. *45*, 1059–1068. https://doi.org/10.1093/nar/gkw1182.

154. Krahn, N., Söll, D., and Vargas-Rodriguez, O. (2022). Diversification of aminoacyl-tRNA synthetase activities via genomic duplication. Front. Physiol. *13*, 983245. https://doi.org/10.3389/fphys.2022.983245.

155. Zhang, C.-M., Liu, C., Slater, S., and Hou, Y.-M. (2008). Aminoacylation of tRNA with phosphoserine for synthesis of cysteinyl-tRNA^{Cys}. Nat. Struct. Mol. Biol. *15*, 507–514. https://doi.org/10.1038/nsmb.1423.

156. Lang, K., and Chin, J.W. (2014). Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. Chem. Rev. *114*, 4764–4806. https://doi.org/10.1021/cr400355w.

157. Kiick, K.L., Saxon, E., Tirrell, D.A., and Bertozzi, C.R. (2002). Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. Proc. Natl. Acad. Sci. USA *99*, 19–24. https://doi.org/10.1073/pnas.012583299.

158. Kramer, G., Sprenger, R.R., Nessen, M.A., Roseboom, W., Speijer, D., de Jong, L., de Mattos, M.J.T., Back, J., and de Koster, C.G. (2010). Proteome-wide alterations in *Escherichia coli* translation rates upon anaerobiosis. Mol. Cell. Proteomics *9*, 2508–2516. https://doi.org/10.1074/mcp.M110.001826.

159. Wiltschi, B., Wenger, W., Nehring, S., and Budisa, N. (2008). Expanding the genetic code of *Saccharomyces cerevisiae* with methionine analogues. Yeast *25*, 775–786. https://doi.org/10.1002/yea.1632.

160. Hartman, M.C.T., Josephson, K., and Szostak, J.W. (2006). Enzymatic aminoacylation of tRNA with unnatural amino acids. Proc. Natl. Acad. Sci. USA *103*, 4356–4361. https://doi.org/10.1073/pnas.0509219103.

161. de la Torre, D., and Chin, J.W. (2021). Reprogramming the genetic code. Nat. Rev. Genet. *22*, 169–184. https://doi.org/10.1038/s41576-020-00307-7.

162. Ros, E., Torres, A.G., and Ribas de Pouplana, L. (2021). Learning from nature to expand the genetic code. Trends Biotechnol. *39*, 460–473. https://doi.org/10.1016/j.tibtech.2020.08.003.

163. Fuertes, G., Sakamoto, K., and Budisa, N. (2023). Exploring and expanding the protein universe with non-canonical amino acids. Front. Mol. Biosci. *10*, 1303286. https://doi.org/10.3389/fmolb.2023.1303286.

164. Fredens, J., Wang, K., de la Torre, D., Funke, L.F.H., Robertson, W.E., Christova, Y., Chia, T., Schmied, W.H., Dunkelmann, D.L., Beránek, V., *et al*. (2019). Total synthesis of *Escherichia coli* with a recoded genome. Nature *569*, 514–518. https://doi.org/10.1038/s41586-019-1192-5.

165. Grome, M.W., Nguyen, M.T.A., Moonan, D.W., Mohler, K., Gurara, K., Wang, S., Hemez, C., Stenton, B.J., Cao, Y., Radford, F., *et al*. (2025). Engineering a genomically recoded organism with one stop codon. Nature *639*, 512–521. https://doi.org/10.1038/s41586-024-08501-x.

166. Sanders, J., Hoffmann, S.A., Green, A.P., and Cai, Y. (2022). New opportunities for genetic code expansion in synthetic yeast. Curr. Opin. Biotechnol. *75*, 102691. https://doi.org/10.1016/j.copbio.2022.102691.

167. Mordaka, P.M., Clouston, K., Cui, J., Holzer, A., Jackson, H.O., Purton, S., and Smith, A.G. (2025). *Chlamydomonas* chloroplast genes tolerate compression of the genetic code to just 51 codons. Preprint at bioRxiv, https://doi.org/10.1101/2025.03.09.641718.

168. Anderson, J.C., Wu, N., Santoro, S.W., Lakshman, V., King, D.S., and Schultz, P.G. (2004). An expanded genetic code with a functional quadruplet codon. Proc. Natl. Acad. Sci. USA *101*, 7566–7571. https://doi.org/10.1073/pnas.0401517101.

169. Xi, Z., Davis, L., Baxter, K., Tynan, A., Goutou, A., and Greiss, S. (2021). Using a quadruplet codon to expand the genetic code of an animal. Nucleic Acids Res. *50*, 4801–4812. https://doi.org/10.1093/nar/gkab1168.

170. Pigula, M.L., Ban, Y., and Schultz, P.G. (2024). Toward a quadruplet codon mitochondrial genetic code. ACS Synth. Biol. *13*, 4175–4179. https://doi.org/10.1093/nar/gkab1168.

171. Fischer, E.C., Hashimoto, K., Zhang, Y., Feldman, A.W., Dien, V.T., Karadeema, R.J., Adhikary, R., Ledbetter, M.P., Krishnamurthy, R., and Romesberg, F.E. (2020). New codons for efficient production of unnatural proteins in a semisynthetic organism. Nat. Chem. Biol. *16*, 570–576. https://doi.org/10.1038/s41589-020-0507-z.

172. Liu, J., Yan, X., Wu, H., Ji, Z., Shan, Y., Wang, X., Ran, Y., Ma, Y., Li, C., Zhu, Y., *et al*. (2025). RNA codon expansion via programmable pseudouridine editing and decoding. Nature *643*, 1410–1420. https://doi.org/10.1038/s41586-025-09165-x.

173. Dunkelmann, D.L., Piedrafita, C., Dickson, A., Liu, K.C., Elliott, T.S., Fiedler, M., Bellini, D., Zhou, A., Cervettini, D., and Chin, J.W. (2024). Adding α,α-disubstituted and β-linked monomers to the genetic code of an organism. Nature *625*, 603–610. https://doi.org/10.1038/s41586-023-06897-6.

174. Cohen, Y., and Alfonta, L. (2025). Engineering of the genetic code. Curr. Opin. Biotechnol. *91*, 103245. https://doi.org/10.1016/j.copbio.2024.103245.

175. Maini, R., Chowdhury, S.R., Dedkova, L.M., Roy, B., Daskalova, S.M., Paul, R., Chen, S., and Hecht, S.M. (2015). Protein synthesis with ribosomes selected for the incorporation of β-amino acids. Biochemistry *54*, 3694–3706. https://doi.org/10.1021/acs.biochem.5b00389.

176. Ohta, A., Murakami, H., Higashimura, E., and Suga, H. (2007). Synthesis of polyester by means of genetic code reprogramming. Chem. Biol. *14*, 1315–1322. https://doi.org/10.1016/j.chembiol.2007.10.015.

177. Guo, J., Wang, J., Anderson, J.C., and Schultz, P.G. (2008). Addition of an α-hydroxy acid to the genetic code of bacteria. Angew. Chem. Int. Ed. Engl. *47*, 722–725. https://doi.org/10.1002/anie.200704074.

178. Sigal, M., Matsumoto, S., Beattie, A., Katoh, T., and Suga, H. (2024). Engineering tRNAs for the ribosomal translation of non-proteinogenic monomers. Chem. Rev. *124*, 6444–6500. https://doi.org/10.1021/acs.chemrev.3c00894.

179. Rezhdo, A., Islam, M., Huang, M., and Van Deventer, J.A. (2019). Future prospects for noncanonical amino acids in biological therapeutics. Curr. Opin. Biotechnol. *60*, 168–178. https://doi.org/10.1016/j.copbio.2019.02.020.

180. Chang, T., Ding, W., Yan, S., Wang, Y., Zhang, H., Zhang, Y., Ping, Z., Zhang, H., Huang, Y., Zhang, J., *et al*. (2023). A robust yeast biocontainment system with two-layered regulation switch dependent on unnatural amino acid. Nat. Commun. *14*, 6487. https://doi.org/10.1038/s41467-023-42358-4.

181. Beattie, A.T., Dunkelmann, D.L., and Chin, J.W. (2023). Quintuply orthogonal pyrrolysyl-tRNA synthetase/tRNA(Pyl) pairs. Nat. Chem. *15*, 948–959. https://doi.org/10.1038/s41557-023-01232-y.

182. Niu, W., and Guo, J. (2024). Cellular site-specific incorporation of noncanonical amino acids in synthetic biology. Chem. Rev. *124*, 10577–10617. https://doi.org/10.1021/acs.chemrev.3c00938.

183. Jann, C., Giofré, S., Bhattacharjee, R., and Lemke, E.A. (2024). Cracking the code: reprogramming the genetic script in prokaryotes and eukaryotes to harness the power of noncanonical amino acids. Chem. Rev. *124*, 10281–10362. https://doi.org/10.1021/acs.chemrev.3c00878.

184. Melnikov, S.V., and Söll, D. (2019). Aminoacyl-tRNA synthetases and tRNAs for an expanded genetic code: what makes them orthogonal? Int. J. Mol. Sci. *20*, 1929. https://doi.org/10.3390/ijms20081929.

185. Dunkelmann, D.L., and Chin, J.W. (2024). Engineering pyrrolysine systems for genetic code expansion and reprogramming. Chem. Rev. *124*, 11008–11062. https://doi.org/10.1021/acs.chemrev.4c00243.

186. Willis, J.C.W., and Chin, J.W. (2018). Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. Nat. Chem. *10*, 831–837. https://doi.org/10.1038/s41557-018-0052-5.

187. Zhang, Q., Jiang, L., Niu, Y., Li, Y., Chen, W., Cheng, J., Ding, H., Chen, B., Liu, K., Cao, J., *et al*. (2025). Machine learning-guided evolution of pyrrolysyl-tRNA synthetase for improved incorporation efficiency of diverse noncanonical amino acids. Nat. Commun. *16*, 6648. https://doi.org/10.1038/s41467-025-61952-2.

188. Reinkemeier, C.D., Girona, G.E., and Lemke, E.A. (2019). Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. Science *363*, eaaw2644. https://doi.org/10.1126/science.aaw2644.

189. Schmied, W.H., Tnimov, Z., Uttamapinant, C., Rae, C.D., Fried, S.D., and Chin, J.W. (2018). Controlling orthogonal ribosome subunit interactions enables evolution of new function. Nature *564*, 444–448. https://doi.org/10.1038/s41586-018-0773-z.

190. Ishida, S., Ngo, P.H.T., Gundlach, A., and Ellington, A. (2024). Engineering ribosomal machinery for noncanonical amino acid incorporation. Chem. Rev. *124*, 7712–7730. https://doi.org/10.1021/acs.chemrev.3c00912.

191. Prokopchuk, G., Butenko, A., Dacks, J.B., Speijer, D., Field, M.C., and Lukeš, J. (2023). Lessons from the deep: mechanisms behind diversification of eukaryotic protein complexes. Biol. Rev. *98*, 1910–1927. https://doi.org/10.1111/brv.12988.

192. Barcytė, D., Žihala, D., and Eliáš, M. (2025). Expanded diversity of pedinophytes provides a window into the evolution of the genetic code in organelles. PLoS Genet. *21*, e1011901. https://doi.org/10.1371/journal.pgen.1011901.

193. Vasquez, Y.M., Nardi, T., Terasaki, G.M., Byl, P.K., Bruna, T., Villada, J.C., Romero, M.F., Mock, T., James, T., GVMAGs data consortium, *et al*. (2025). Genome-resolved expansion of Nucleocytoviricota and Mirusviricota reveals new diversity, functional potential, and biotechnological applications. Preprint at bioRxiv, https://doi.org/10.1101/2025.09.26.678796.

194. Robertson, W.E., Rehm, F.B.H., Spinck, M., Schumann, R.L., Tian, R., Liu, W., Gu, Y., Kleefeldt, A.A., Day, C.F., Liu, K.C., *et al*. (2025). *Escherichia coli* with a 57-codon genetic code. Science *390*, eady4368. https://doi.org/10.1126/science.ady4368.

195. Fournier, G.P., Huang, J., and Gogarten, J.P. (2009). Horizontal gene transfer from extinct and extant lineages: biological innovation and the coral of life. Philos. Trans. R. Soc. Lond. B Biol. Sci. *364*, 2229–2239. https://doi.org/10.1098/rstb.2009.0033.

196. Mukai, T., Amikura, K., Fu, X., Söll, D., and Crnković, A. (2022). Indirect routes to aminoacyl-tRNA: the diversity of prokaryotic cysteine encoding systems. Front. Genet. *12*, 794509. https://doi.org/10.3389/fgene.2021.794509.

197. Lewis, A.M., Fallon, T., Dittemore, G.A., and Sheppard, K. (2024). Evolution and variation in amide aminoacyl-tRNA synthesis. IUBMB Life *76*, 505–522. https://doi.org/10.1002/iub.2811.

198. Brown, J.R., and Doolittle, W.F. (1999). Gene descent, duplication, and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. J. Mol. Evol. *49*, 485–495. https://doi.org/10.1007/PL00006571.

199. Numata, T. (2015). Mechanisms of the tRNA wobble cytidine modification essential for AUA codon decoding in prokaryotes. Biosci. Biotechnol. Biochem. *79*, 347–353. https://doi.org/10.1080/09168451.2014.975185.

200. Taniguchi, T., Miyauchi, K., Nakane, D., Miyata, M., Muto, A., Nishimura, S., and Suzuki, T. (2013). Decoding system for the AUA codon by tRNA Ile with the UAU anticodon in *Mycoplasma mobile*. Nucleic Acids Res. *41*, 2621–2631. https://doi.org/10.1093/nar/gks1344.

201. van der Gulik, P.T.S., and Hoff, W.D. (2016). Anticodon modifications in the tRNA set of LUCA and the fundamental regularity in the standard genetic code. PLoS One *11*, e0158342. https://doi.org/10.1371/journal.pone.0158342.

202. Wang, X., and Lavrov, D.V. (2011). Gene recruitment — a common mechanism in the evolution of transfer RNA gene families. Gene *475*, 22–29. https://doi.org/10.1016/j.gene.2010.12.009.

203. Sahyoun, A.H., Hölzer, M., Jühling, F., Höner zu Siederdissen, C., Al-Arab, M., Tout, K., Marz, M., Middendorf, M., Stadler, P.F., and Bernt, M. (2015). Towards a comprehensive picture of alloacceptor tRNA remolding in metazoan mitochondrial genomes. Nucleic Acids Res. *43*, 8044–8056. https://doi.org/10.1093/nar/gkv746.

204. Schneider, A. (2011). Mitochondrial tRNA import and its consequences for mitochondrial translation. Annu. Rev. Biochem. *80*, 1033–1053. https://doi.org/10.1146/annurev-biochem-060109-092838.

205. Lang, B.F., Lavrov, D., Beck, N., and Steinberg, S.V. (2012). Mitochondrial tRNA structure, identity, and evolution of the genetic code. In Organelle Genetics: Evolution of Organelle Genomes and Gene Expression, C.E. Bullerwell, ed. (Springer Berlin Heidelberg), pp. 431–474. https://doi.org/10.1007/978-3-642-22380-8_17.

206. Kijima, S., Hikida, H., Delmont, T.O., Gaïa, M., and Ogata, H. (2024). Complex genomes of early nucleocytoviruses revealed by ancient origins of viral aminoacyl-tRNA synthetases. Mol. Biol. Evol. *41*, msae149. https://doi.org/10.1093/molbev/msae149.

207. Costello, A., and Badran, A.H. (2021). Synthetic biological circuits within an orthogonal central dogma. Trends Biotechnol. *39*, 59–71. https://doi.org/10.1016/j.tibtech.2020.05.013.

208. Lee, J.Y., Kim, D.G., Kim, B.-G., Yang, W.S., Hong, J., Kang, T., Oh, Y.S., Kim, K.R., Han, B.W., Hwang, B.J., *et al*. (2014). Promiscuous methionyl-tRNA synthetase mediates adaptive mistranslation to protect cells against oxidative stress. J. Cell Sci. *127*, 4234–4245. https://doi.org/10.1242/jcs.152470.

209. Bezerra, A.R., Guimarães, A.R., and Santos, M.A.S. (2015). Non-standard genetic codes define new concepts for protein engineering. Life *5*, 1610–1628. https://doi.org/10.3390/life5041610.

210. Inagaki, Y., Blouin, C., Doolittle, W.F., and Roger, A.J. (2002). Convergence and constraint in eukaryotic release factor 1 (eRF1) domain 1: the evolution of stop codon specificity. Nucleic Acids Res. *30*, 532–544. https://doi.org/10.1093/nar/30.2.532.

211. Lekomtsev, S., Kolosov, P., Bidou, L., Frolova, L., Rousset, J.-P., and Kisselev, L. (2007). Different modes of stop codon restriction by the *Stylonychia* and *Paramecium* eRF1 translation termination factors. Proc. Natl. Acad. Sci. USA *104*, 10824–10829. https://doi.org/10.1073/pnas.0703887104.

212. Eliseev, B., Kryuchkova, P., Alkalaeva, E., and Frolova, L. (2010). A single amino acid change of translation termination factor eRF1 switches between bipotent and omnipotent stop-codon specificity. Nucleic Acids Res. *39*, 599–608. https://doi.org/10.1093/nar/gkq759.

213. Lind, C., Oliveira, A., and Åqvist, J. (2017). Origin of the omnipotence of eukaryotic release factor 1. Nat. Commun. *8*, 1425. https://doi.org/10.1038/s41467-017-01757-0.

214. Nagao, A., Ohara, M., Miyauchi, K., Yokobori, S.-i., Yamagishi, A., Watanabe, K., and Suzuki, T. (2017). Hydroxylation of a conserved tRNA modification establishes non-universal genetic code in echinoderm mitochondria. Nat. Struct. Mol. Biol. *24*, 778–782. https://doi.org/10.1038/nsmb.3449.

215. Lavrov, D.V., and Pett, W. (2016). Animal mitochondrial DNA as we do not know it: mt-genome organization and evolution in nonbilaterian lineages. Genome Biol. Evol. *8*, 2896–2913. https://doi.org/10.1093/gbe/evw195.

216. Alfonzo, J.D., Blanc, V., Estévez, A.M., Rubio, M.A.T., and Simpson, L. (1999). C to U editing of the anticodon of imported mitochondrial $tRNA^{Trp}$ allows decoding of the UGA stop codon in *Leishmania tarentolae*. EMBO J. *18*, 7056–7062. https://doi.org/10.1093/emboj/18.24.7056.

217. Paris, Z., Svobodová, M., Kachale, A., Horáková, E., Nenarokova, A., and Lukeš, J. (2021). A mitochondrial cytidine deaminase is responsible for C to U editing of $tRNA^{Trp}$ to decode the UGA codon in *Trypanosoma brucei*. RNA Biol. *18*, 278–286. https://doi.org/10.1080/15476286.2021.1940445.

218. Swart, E.C., Emmerich, C., Seah, K.B.B., Singh, M., Shulgina, Y., and Singh, A. (2024). How did UGA codon translation as tryptophan evolve in certain ciliates? A critique of Kachale *et al.* 2023 Nature. eLife *13*, RP93502. https://doi.org/10.7554/elife.93502.

219. Gissi, C., Iannelli, F., and Pesole, G. (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity *101*, 301–320. https://doi.org/10.1038/hdy.2008.62.

220. Ohira, T., Suzuki, T., Miyauchi, K., Suzuki, T., Yokobori, S.-i., Yamagishi, A., and Watanabe, K. (2013). Decoding mechanism of non-universal genetic codes in *Loligo bleekeri* mitochondria. J. Biol. Chem. *288*, 7645–7652. https://doi.org/10.1074/jbc.M112.439554.

221. Kollmar, M., and Mühlhausen, S. (2017). How tRNAs dictate nuclear codon reassignments: Only a few can capture non-cognate codons. RNA Biol. *14*, 293–299. https://doi.org/10.1080/15476286.2017.1279785.

222. Su, D., Lieberman, A., Lang, B.F., Simonović, M., Söll, D., and Ling, J. (2011). An unusual $tRNA^{Thr}$ derived from $tRNA^{His}$ reassigns in yeast mitochondria the CUN codons to threonine. Nucleic Acids Res. *39*, 4866–4874. https://doi.org/10.1093/nar/gkr073.

223. Holman, K.M., Wu, J., Ling, J., and Simonović, M. (2015). The crystal structure of yeast mitochondrial ThrRS in complex with the canonical threonine tRNA. Nucleic Acids Res. *44*, 1428–1439. https://doi.org/10.1093/nar/gkv1501.