

The Unexpected Strategies behind Alternative Genetic Codes

Jef Struyf*

A Retired Lecturer of UCLL Campus Gasthuisberg, Chemistry Bachelors Herestraat 49, Be-3000 Leuven, Belgium

*Corresponding author: jef.struyf@ucll.be

Received May 19, 2021; Revised June 23, 2021; Accepted July 02, 2021

Abstract This article proposes a number of codon reassignment strategies. These strategies are inferred by comparing the alternative (non-standard) genetic codes to the standard genetic code. Codon reassignments show strong and non-random connection to the standard code. The strategies interconnect all genetic codes and model a selected pathway through the heterotrophic part of life. The article also shows that the codon reassignment strategies model the evolution from the standard to the vertebrate mitochondrial code. Furthermore, it introduces a genetic code table based on third position synonyms that compares the vertebrate mitochondrial code to the standard code.

Keywords: *alternative genetic codes, biochemical strategies, codon reassignments, codon reassignment strategies, codon reassignment tables, genetic code table, molecular evolution, molecular genetics, non-standard genetic codes, table for alternative genetic codes, table for non-standard genetic codes*

Cite This Article: Jef Struyf, “The Unexpected Strategies behind Alternative Genetic Codes.” *American Journal of Educational Research*, vol. 9, no. 7 (2021): 417-425. doi: 10.12691/education-9-7-4.

1. Introduction

The genetic code is the key of the chemistry of life and almost universal. Altogether, the genetic codes (standard and non-standard) show molecular results of the evolutionary process, which is the most reliable and consistent experiment ever executed. We consider the following synonyms: “The alternative genetic codes” and “the non-standard genetic codes”. The codons in this contribution are mRNA codons. Reference 1 (Open Access), Section 1.1 “The Chemistry of Life”, can be consulted by readers who have insufficient knowledge of biochemistry to get a high-level introduction [1].

1.1. Codon Reassignments Data and General Characteristics

The deciphering of the genetic code determines the standard amino acids and the standard genetic code [2]. A codon reassignment assigns (encodes) a different standard amino acid than the standard genetic code assignment. For some organisms, some codons (maximum six for one species, family, order, class or subphylum) reassign (encode) other standard amino acids than the standard code. The National Center for Biotechnology Information (NCBI) compiles organisms that have codon reassignments and compare them to the standard genetic code [3]. A superficial look at the NCBI list of non-standard genetic codes shows many but often identical codon reassignments spread over twenty-four genetic codes or translation tables (TTs) not including the

standard code (TT 1). The 24 non-standard TTs are not contradictory to the fact that for example TT 33 exists. NCBI omitted some TTs from the list probably due to unconfirmed results.

Figure 1 shows the TTs of eumetazoa and Figure 2 shows those of unicellulars. The list of alternative genetic codes includes following organisms: One prokaryote, fourteen eukaryote unicellulars, and nine eumetazoa. Eumetazoa are a subkingdom of the animalia kingdom. Eumetazoa include all the animals except the sponges. The ten reassignment sensitive eumetazoa are: Fungi, various non-vertebrates and the vertebrates. All the eumetazoan alternative codes are mitochondrial. Unicellulars show nine nuclear and five mitochondrial TTs. The TT list is rather limited. Evolution easily could have produced much more deviations from the standard code because alternative genetic codes don't influence the food chains. They don't change the standard amino acid contingent. The NCBI reference includes the original references and the systematic range of the organisms for each TT. The 25 TTs compiled at NCBI count only thirteen (13) different reassignment sensitive codons out of the sixty-four (64) canonical codons. Unicellular organisms count ten (10) sensitive codons: The UUA, UCA, UAR (two codons), UGA, CUN (four codons), and AUA codons. The eumetazoa count six sensitive codons: The UAA*, UGA*, AUA*, AAA, and AGR (two codons) codons. The vertebrates count only four sensitive codons: UGA*, AUA*, and AGR (two codons). An asterisk marks the codons that are in common with unicellulars. Reference 4 shows nucleotide and nucleotide combination symbols [4]. Frequently used examples are: A or G or C or U is N, A or G is R, C or U is Y, Y or A is H and Y or G is B.

Third Position Synonyms for the Standard Code	UAR		AUH	AAR	AGR	AGR
Sensitive Codons for Reassignments	UAA	UGA	AUA	AAA	AGA	AGG
Standard, Universal Code (TT 1)	Ter	Ter	Ile	Lys	Arg	Arg

Mitochondrial Codes for Eumetazoa	Codon Reassignments					
Plant (TT 1)						
Mold (Multicellular Fungi), (TT 4)		Trp				
Coelenterate (TT 4), (Cnidaria, Ctenophora)		Trp				
Rhabdopleuridae, Pterobranchia (TT 24), (Hemichordata)		Trp			Ser	Lys
Cephalodiscidae (TT 33), (Hemichordata)	Tyr	Trp			Ser	Lys
Alternative Flatworm (TT 14), (Platyhelminthes)	Tyr	Trp		Asn	Ser	Ser
Echinoderm and Flatworm (TT 9)		Trp		Asn	Ser	Ser
Trematode (TT 21), (Platyhelminthes)		Trp	Met	Asn	Ser	Ser
Invertebrate (TT 5) (Arthropoda, Mollusca, Nematoda)		Trp	Met		Ser	Ser
Ascidian (TT 13), (Chordata, Tunicata)		Trp	Met		Gly	Gly
Vertebrate (TT 2)		Trp	Met		Ter	Ter

The smaller upper part (three rows) of Figure 1 shows the standard codon assignment (row 3) for the reassignment sensitive codons (row 2) and the standard third position synonyms (row 1). The larger lower part of Figure 1 shows the codon reassignments for eumetazoans. Three letter abbreviations show the (re)assigned amino acids. Ter is a termination or stop codon (re)assignment. A blank field represents a standard assignment. R is A + G, Y is U + C and H is Y + A. Following reassignments determine the eumetazoan sequence: The polarity sequence, sequence overlap 1 reassignments, sequence overlap 2 reassignments, sequence overlap 3 reassignments, sequence overlap 4 reassignments, sequence overlap 5 reassignments. The data are from “The National Center for Biotechnology Information” (NCBI).

Figure 1. Eumetazoan mitochondrial reassignments compared to the standard genetic code assignments

Third Position Synonyms for the Standard Code	UUR	UCN	UAR	UAR		CUN	AUH
Sensitive Codons for Reassignments	UUA	UCA	UAR	UAG	UGA	CUN, (CUG)	AUA
Standard, Universal Code (TT 1)	Leu	Ser	Ter	Ter	Ter	Leu	Ile

Genetic Codes	Codon Reassignments						
Candidate Div. SR1, Gracilibacteria (TT25)							Gly
Nuclear Codes							
Mostly Ciliates and Flagellates							
Euplotidae (TT 10) a ciliate							Cys
Condylostoma (TT 28) a ciliate				Ter/Gln			Ter/Trp
Ciliate, dasycladacean and hexamita (TT 6)				Gln			
Karyorelict (TT 27) a ciliate				Gln			Ter/Trp
Blastocriothidia (TT 31) a flagellate				Ter/Glu			Trp
Peritrich (TT 30) a ciliate				Glu			
Mesodinium (TT 29) a ciliate				Tyr			
Saccharomyces							
Alternative Yeast (TT 12)							Ser (CUG)
Pachysolen tannophilus (TT 26) a yeast							Ala (CUG)
Mitochondrial Codes							
Protozoan (TT 4)							Trp
Yeast (TT 3), (Saccharomyces)							Trp
Some Algae							
Chlorophycean (TT 16)							Leu
Scenedesmus Obliquus (TT 22)							Leu
Thraustochytrium aureum (TT 23)							Ter

The smaller upper part (three rows) of Figure 1 shows the standard codon assignment (row 3) for the reassignment sensitive codons (row 2) and the standard third position synonyms (row 1). The larger lower part of Figure 1 shows the codon reassignments for eumetazoans. Three letter abbreviations show the (re)assigned amino acids. Ter is a termination or stop codon (re)assignment. A blank field represents a standard assignment. R is A + G, Y is U + C and H is Y + A. Three color high lightings (red, turquoise and yellow) mark some of the connections to Figure 1: The glycine-cysteine-tryptophan polarity sequence, the glutamine-glutamic acid-tyrosine balance and the yeast-vertebrate connection. Pink color highlight marks the nuclear-mitochondrial evolution for the Saccharomyces. The data are from “The National Center for Biotechnology Information” (NCBI).

Figure 2. Unicellular codon reassignments compared to the standard genetic code codon assignments

1.2. The Codon Reassignment Research

The reassignment research makes partial comparisons and optimality calculations [5], defines new concepts for protein engineering [6], compares translation loads [7], describes a reassignment model [8], reassigns

non-canonical amino acids [9], discusses the evolution of alternative genetic codes [10-15], shows how mitochondria redefine the code [16], defines an extreme codon preference: Reassignments [17], mechanisms of codon reassignments [18,19] and alternative codes are more robust to amino acid replacements [20]. These

articles explain the codon reassignments as a mutation driven event, a result of non-adaptive evolution, a result of codon disappearance (for smaller genomes) and reassigned reappearance, by ambiguous intermediates, by finding accompanying facts such as the gain or the loss of a tRNA or release factor, and by coding ambiguity, which is the extension of a tRNA's decoding capacity beyond its original set of codons. The explanations for codon reassignment demonstrate the underlying molecular mechanisms.

1.3. Codon Reassignments and Educational Practice

In order to read most research articles, including the above referenced articles on codon reassignments, the reader needs to possess specific skills and domain specific background. Therefore, educators and students consult textbooks. Many textbooks do not cover the subject of codon reassignments. That is why this subject is not so popular. The present article helps to make the codon reassignment subject more accessible. The reading threshold of the present article is low and reading it doesn't require profound subject matter expertise. A science graduate is perfect and with some help of the teacher it is accessible for last year high school students. The article broadens the student's horizon on genetics.

Our investigation reveals non-random codon reassignment strategies, provides the instructor with practical overview tables, and shows an evolutionary pathway between the standard genetic code and the vertebrate mitochondrial code. For educational purposes it is even not necessary to study each strategy in detail. Instead one can select a few strategies for a deeper study to judge their influence, reliability and validity. A strategy name in combination with Figures 1-3 makes most strategies more directly accessible. Please refer to the figures while reading the text. The strategies disprove the general accepted opinion that codon reassignments are random events that only have to be supported by their molecular mechanisms, which is the target of traditional research on the subject.

2. Method

2.1. The Comparison Method

This investigation contains no experimental procedures conducted by the author, who uses the NCBI data of "The Genetic Codes" [3]. The detection of the reassignment strategies requires an adapted method: The comparison method. By comparing the codon reassignments to each other, to the standard code, to neighboring codons, to related translation tables, to terminations (terminations) and by comparing unicellular reassignments to eumetazoan, we find and describe intrinsic determinant strategies. The strategies are logical expressions of the investigated connections. The codon reassignment strategies function at the level of the coding process and determine the non-standard encoded standard amino acids or termination.

2.2. Calculation for Strategy Contribution

We express the contribution of a strategy by the number of concerned reassignments to the total number (18) of different reassignments. Many strategies show overlap: Reassignments that are in common to more than one strategy. The contribution calculations of the strategies with overlapping connections count only the non-overlapping reassignments.

2.3. Description of the Reassignment Overview Tables

Two figures arrange the NCBI codon reassignments. The smaller upper part of [Figure 1](#) and [Figure 2](#) shows the standard codon assignments (row 3) for the reassignment sensitive codons (row 2) and the standard third position synonyms (row 1). The larger lower part of [Figure 1](#) and [Figure 2](#) shows the codon reassigned amino acids by three letter abbreviations for eumetazoan and for unicellular organisms, respectively. The unicellulars are possibly including multicellular aggregates. We don't represent all the reassignments in one table, because such a table would have too many columns and too many blank fields. Blank fields correspond to standard assignments. Together, [Figure 1](#) and [Figure 2](#) count 18 different reassignments including the reassignment repeats at different sensitive codons except the repeats at third position synonyms. [Figure 1](#) and [Figure 2](#) count 5 and 10 figure specific different reassignments, respectively, and have 3 reassignments (UGA tryptophan, AUA methionine and UAA tyrosine) in common to both tables.

3. Results and Discussion

3.1. The Strategies of the Eumetazoan Mitochondrial Codon Reassignments

All non-standard eumetazoan genetic codes are mitochondrial. The comparison of eumetazoan codes to the standard code determines a TT sequence that is supported by the codon reassignments of the corresponding TTs ([Figure 1](#)).

3.1.1. The Construction of the Eumetazoan Sequence by the Polarity Sequence and the Sequence Overlap Strategies

This section describes two reassignment strategies: The polarity sequence strategy and the sequence overlap strategy. [Figure 1](#) constructs the eumetazoan TT sequence by comparing the eumetazoan codon reassignments. The overlapping reassignments may not be interrupted by a standard assignment. The sequence is in between the TT 2 AGR and the TT 1 UGA terminations. From the TT 2 AGR terminations on, the sequence shows a gradual increase in the correspondence of the codon reassignments to the standard code by the polarity increase of the AGG reassignments from glycine, over serine and lysine to the arginine standard (TT 4). That's why we construct the sequence from bottom to top of the larger lower part of [Figure 1](#). A color highlight marks the polarity sequence and the various overlaps. The sequence overlap reassignments

are (see Figure 1): AUA methionine (overlap 1), AAA asparagine (overlap 2), UAA tyrosine (overlap 3), AGA termination - glycine - serine (overlap 4), and the overlap of a number UGA tryptophan reassignments (overlap 5). The AGA codons don't reassign lysine. Compared to the AGG reassignments, the AGA reassignments have an incomplete polarity sequence. That's why the AGA reassignments are useful as overlap reassignments. The AGG polarity sequence (glycine, serine, lysine, arginine) in combination with the AUA methionine and AAA asparagine overlaps determine the first three TTs of the eumetazoan sequence: TT 2, TT 13, and TT 5. Reassignment overlaps AUA methionine, AAA asparagine and UAA tyrosine determine the exact position of the TTs 21, 9 and 14. The turquoise (UAA tyrosine) and the pink (AGA termination - glycine - serine) overlaps determine the TTs 33 and 24 position in the sequence. The UGA tryptophan reassignments mark the end of the TT sequence (TT 4). The coelenterate and mold are both part of TT 4. Their relative position in the sequence is not based on reassignment data. Coelenterates belong, like the organism of the lower ranked TTs in Figure 1, to the animalia kingdom. Based on the reassignment data, the fungi kingdom is in between the animalia and the plant kingdoms. The next TT is TT 1 (the standard code) for plant mitochondria. Coelenterates, molds and plants show the standard AGR arginine assignment. Column 1 of Figure 1 shows the eumetazoan sequence. Any other TT sequence, constructed with the investigated NCBI data, violates at least one overlap sequence, which means that at least one overlap sequence will be interrupted by a standard assignment. The polarity sequence strategy and the sequence overlaps count for 22.2 % (4/18) and 27.8 (5/18) of the different reassignments, respectively.

3.1.2. The Lysine Codon Degeneracy Compensatory Strategy for Echinodermata and Hemichordata

The lysine codon degeneracy compensation is a connection between Hemichordata TTs 24 and 33, and Echinoderm and Platyhelminthes TTs 9, 14 and 21. For the Hemichordata of TTs 24 and 33, the AGG lysine together with the standard AAR lysine result in a triplet degeneracy (three synonymous codons). In contrast, the TTs 9, 14 and 21 show the singlet AAG lysine because AAA is reassigned into asparagine. The Hemichordata (TTs 24 and 33) overcompensate the lack of lysine codons for TTs 9, 14 and 21. The standard code has a doublet lysine degeneracy and is therefore the equilibrium (balance) point of the compensatory strategy. Echinodermata (TT 9) and Hemichordata are sister groups [21], which makes the compensation more logical. In summary, the compensation consists of:

The singlet AAG lysine for TTs 9, 14, 21, the doublet AAR (two codons) lysine for the standard code and the triplet AAR - AAG lysine for TTs 24 and 33. The compensation extends the Hemichordata / Echinodermata connection of reference 13 to the Platyhelminthes. This compensation strategy counts for 11.1 % (2/18) of the different reassignments and connects to the standard code.

3.1.3. The Eumetazoan Unit Structure

The AGG polarity sequence and the assisting overlap codon reassignments, are all essential and precisely the

necessary ones for constructing the eumetazoan TT sequence. Not a single eumetazoan reassignment is out of the scope of this remarkable TT sequence that possibly forms evidence for the evolution of TT 1 to TT 2 code or vice versa. In any case, the TT sequence challenges the scientific community for an explanation. All the eumetazoan reassignments form a network (the eumetazoan unit structure including the corresponding TT sequence) and connect via the AGG polarity sequence to the standard genetic code. The eumetazoan unit structure counts for 44.4 % (8/18) of the different reassignments.

3.1.4. The Termination Codon Degeneracy of Eumetazoan Mitochondria

The eumetazoan mitochondria have four different termination codon degeneracies depending on the involved TTs (Figure 1). These four degeneracies are as follows:

Singlet degeneracy by the UAG termination of TTs 14 and 33.

Doublet degeneracy by the UAR terminations of non-vertebrate eumetazoa, except TTs 1, 14 and 33.

Triplet degeneracy by the UAR and UGA terminations of plants (TT 1, the standard code).

Quartet degeneracy by the UAR and AGR Terminations of vertebrates (TT 2).

3.2. Codon Reassignment Strategies for Unicellular Organisms

3.2.1. The Codon Reassignment Balance Strategy of Nuclear UAR Codon Reassignments for (mostly) Ciliates

The nuclear UAR reassignments (Figure 2) are glutamic acid, glutamine and tyrosine. The corresponding standard codons are GAR, CAR and UAY, respectively. UAR codons differ from GAR and CAR codons (standard glutamic acid and glutamine) by the first and from UAY (standard tyrosine) by the third nucleotide.

A twofold difference of the UAR codons' first nucleotide (from U to G and C) balances a twofold difference of the UAR codons' third nucleotide: R versus Y. The standard UAR termination codons form the equilibrium (balance) point. The codon reassignment balance strategy interconnects the UAR reassignments of ciliates and flagellates. Thanks to this balance strategy, the TT 29 UAA tyrosine reassignment connects to the eumetazoan UAA tyrosine of TTs 14 and 33. The codon reassignment balance strategy counts 16.7% (3/18) of the different reassignments.

3.2.2. A Connecting Strategy between Unicellular Nuclear and Mitochondrial Genetic Codes to the Mitochondrial Eumetazoan Codes

The comparison of TTs for unicellular organisms connects nuclear to mitochondrial codes, including mitochondrial TTs for eumetazoan organisms. The following four examples, representing most unicellular reassignments, show the connection.

The UGA Tryptophan Connection

Firstly, the nuclear TTs 27, 28 and 31, and mitochondrial TTs 3 and 4 codes are in line with the

corresponding mitochondrial eumetazoan codes for heterotrophs because they all reassign the standard UGA termination to UGA tryptophan.

The UAA Tyrosine Connection

Secondly, the TT 29 UAA tyrosine of the ciliate *Mesodinium* connects to eumetazoan TTs 14 and 33 UAA tyrosine.

The Saccharomyces Two-steps Connection

Thirdly, the *Saccharomyces* show a two-step connection to the eumetazoan sequence. Step 1, the nuclear reassignments for TTs 12 and 26 (serine and alanine, respectively) are at the CUG codon, which according to codon structure are in line with the CUN threonine of the mitochondrial TT 3 (Section 3.2.3.). Step 2, the AUA methionine of the mitochondrial TT 3 is identical to the AUA methionine of four eumetazoan TTs: 2, 5, 13 and 21, and the TT 3 UGA tryptophan is already mentioned in **The UGA Tryptophan Connection** of this Section 3.2.2.

Comparable Bridging Sequences

Fourthly, unicellular nuclear UGA and eumetazoan AGG reassignments show comparable bridging sequences. We compare the sequence of nuclear UGA (re)assignments from [Figure 2](#): Termination (the standard), glycine, cysteine and tryptophan for unicellulars with the sequence of mitochondrial eumetazoan AGG (re)assignments from [Figure 1](#): termination, glycine, serine, lysine and standard arginine for eumetazoa. Both sequences have the first two (re)assignments in common (termination and glycine). For the next in line of the sequences, the cysteine/serine reassignments, cysteine (for the UGA sequence) is a sulfur derivative of serine (for the AGG sequence). Cysteine and serine are best in line with the increase, respectively weaker and stronger, of the side chain polarity increase over their respective sequence. The eumetazoan sequence needs an additional reassignment (AGG lysine) to bridge the polarity jump towards arginine of TT 1. The eumetazoan AGG bridging sequence ends at the TT 4 AGG arginine, which is the standard assignment. The AGG bridging sequence is interrupted by a number of serine reassignments and the UGA sequence ends in a longer number of tryptophan reassignments from unicellular nuclear over unicellular mitochondrial extending into the eumetazoan TT sequence and including TT 2 UGA tryptophan. The connection between both bridging sequences reveals a comparable strategy for unicellular as for eumetazoan codes. **Just like the eumetazoa, the unicellulars show a polarity sequence strategy.** The connection for both sequences is a pearl of a connecting strategy. Both sequences bridge TT 2 to TT 1. We assume that all mitochondrial codes for heterotrophs and some nuclear codes for unicellulars reassign UGA from termination (standard) to tryptophan. **The four items of Section 3.2.2. prove that unicellular evolution tends to reach mitochondrial eumetazoan codon reassignments.** The four connecting examples count for 55.6 % (10/18) of the different reassignments and 76.9 % (10/13) of the different unicellular reassignments. Only the algae are not yet involved, but they will be by Section 3.2.5.

3.2.3. The Strategy of *Saccharomyces* CUG and CUN Codon Reassignments

The *Saccharomyces* nuclear CUG (TT 12 and 26) and mitochondrial CUN (TT 3) reassignments are serine, alanine and threonine, respectively. The standard codons for alanine, serine and threonine (GCN, UCN and ACN) reciprocally differ only in the first position (are first position neighbors). Standard CCN proline is also a first position neighbor, but because proline causes a nick in the protein chain, proline is not a favorable candidate for a reassignment. Standard GUN valine is a first position neighbor of the standard CUN leucine. The polarity correspondence between leucine and valine makes valine a very acceptable reassignment candidate. Notwithstanding the latter fact, *Saccharomyces* prefer acceptable amino acid reassignments (not proline) of which the standard codons are first position neighbors. The *Saccharomyces* strategy connects strongly to the standard code and explains the *Saccharomyces* reassignment choices and counts 16.7 % (3/18) of the different reassignments.

3.2.4. The Termination Degeneracy of Unicellulars

The unicellulars have four different termination degeneracies dependent on the involved TT ([Figure 2](#)). These four degeneracies are as follows:

Singlet degeneracy by the UAG termination of TTs 6, 27, 29 and 30.

Doublet degeneracy by the UAR terminations of TTs 3, 4, 10, 25 and 31 and the URA (UGA and UAA) terminations of TT 16.

Triplet degeneracy by the UAR and UGA terminations of TT 28 (like the standard code) and the UVA (V = R + C) terminations of TT 22.

Quartet degeneracy by the TT 23 UKA-UAR terminations (UKA = UGA + UUA).

The singlet, doublet, triplet and quartet degeneracies for unicellular terminations correspond to the degeneracies of eumetazoan terminations.

3.2.5. The Termination Degeneracy Compensatory Strategy (Unicellular Algae)

The Description of the Compensation

This strategy concerns mitochondrial codon reassignments for some unicellular algae of TTs 16, 22 and 23 from [Figure 2](#). Section 3.2.4 shows that the termination degeneracies for the latter TTs are doublet (TT 16), triplet (TT 22) and quartet (TT 23).

More in detail, the TT 23 additional termination is the result of a termination degeneracy compensation strategy between the algae TTs 16, 22 and 23. TT 16 reassigns the standard UAG termination to leucine (a termination loss for TT 16), TT 22 additionally reassigns UCA to a termination thereby restoring the termination degeneracy triplet and TT 23 produces the extra termination by reassigning UUA to termination. The TT 22 termination is the equilibrium (balance) point, because its triplet termination degeneracy is in between TT 16 (a termination loss) and TT 23 (an additional termination). The TT 22 UCA termination replaces the standard UAG termination that becomes UAG leucine.

Eumetazoan Connection and Contributions

The TT 23 additional UUA termination resembles the additional vertebrate TT 2 terminations at AGR. The termination degeneracy compensation resembles the already presented lysine degeneracy compensation strategy for related eumetazoan TTs. Consequently, the similarity of unicellular algae to eumetazoan reassignments connects the reassignments for algae to the eumetazoan unit structure. The termination degeneracy compensation counts 16.7 % (3/18) of the different reassignments of which 23.1 % (3/13) of the different unicellular reassignments.

Phototroph Algae Deviate Heterotroph Reassignments

Notwithstanding the similarities, the phototroph algae deviate heterotroph reassignments by second position synonyms and by the triplet UCB serine third position synonyms. The algae second position termination synonyms are TT 16 URA, TT 23 UKA and TT 22 UVA. TTs 16 and 22 UAG leucine forms with standard UUG leucine the UWG (W = U + A) leucine second position synonyms. The second position synonyms for these algae are somewhat deviating. Most synonyms of non-standard reassignments (and standard assignments) are on the third position. There are a few first or second position synonyms, but these are also third position synonyms. Examples of these non-standard synonyms from Figure 3 are: YAR glutamine and KAR glutamic acid (K is G or U). It is well-known that the biochemistry of phototropic organisms extends the borders of the biochemistry of heterotrophs (e.g. the photosynthesis). Similarly, the non-standard genetic codes of unicellular phototrophs extend the codon reassignments of non-standard heterotrophic organisms with atypical second position synonyms. The TT 22 UCB (B = Y + G) serine synonyms (complemental to the UCA termination) support this statement.

3.3. Common Reassignment Strategies for Unicellulars and Eumetazoa

3.3.1. The Termination Strategy

The TT 1 terminations are at UAR and UGA codons. Termination reassignments are at AGR (TT 2), UCA (TT 22) and UUA (TT23) codons. The TTs 1 and 2 terminations trigger (additional) reassignments (see Figures 2 and 1, respectively). Most reassignments occur at the just mentioned termination positions or are terminations [22]. Only the reassignments for Saccharomyces and the reassignments for the eumetazoan AUA and AAA codons deviate from this rule. The terminations of algae produce no additional reassignments. The termination strategy (at TTs 1 and 2 terminations, and the Algae terminations) counts for 72.2% (13/18) of the different reassignments of which 38.9 % (7/18) to the standard terminations.

3.3.2. The Strategy of the Reassignment Triplets

The reassignments at the various sensitive codons, except the sequence overlap reassignments in Figure 1, tend to form a triplet of three differently encoded amino acids. Figure 1 only shows the AGG glycine/serine/lysine triplet. The other sensitive codons of Figure 1 reassign amino acids that form sequence overlaps. Figure 2 shows

three triplets: the UAR glutamine, glutamic acid and tyrosine triplet, the UGA glycine, cysteine, and tryptophan triplet and the CUN serine, alanine and threonine triplet. As usual, the algae also try to deviate from this strategy but nevertheless show the triplet of UUA termination, UCA termination and UAG leucine. The triplets count 66.7 % (12/18) and including the algae triplet 83.3 % (15/18) of the different reassignments. Furthermore, many reassignments form triplet third position synonyms: UAH tyrosine (TTs 14 and 33), AAH asparagine (TTs 9, 14 and 21), AGH serine (TTs 24 and 33), UGH cysteine (TT 10) and CUH leucine (TTs 12 and 26) and TT 22 UCB serine.

3.3.3. The Neighboring (Adjacent) Codon Strategy

Most codon reassignments show a strong connection/relation to the standard code. Many reassignments (Figure 3, column 2) are the same as the standard assignment of one or more neighboring codons (column 3). We call this the neighboring codon strategy. The connection is less clear for glycine and lysine. A codon neighbor has two codon positions (usually the positions 1 and 2 of the codon triplet) in common. Figure 3 contains 12 different reassignments of which 10 use the standard assignment of neighboring codons. All the reassignments of Figure 1 are also part of Figure 3. In addition to the eumetazoan strategies discussed in Section 3.1., the reassignments of Figure 1 follow the neighboring codon strategy. The neighboring codon strategy connects 55.6 % (10/18) of the different reassignments to the standard code. Without glycine and lysine this becomes 44.4 % (8/18).

3.3.4. The Strategy of the Third Position Synonyms

Most codon synonyms are third position synonyms. The standard code counts four: N, R, Y and H (AUH isoleucine) of the eleven possible nucleotide combinations for third position synonyms (R, Y, S, W, K, M, B, D, H, V, N from SMS-IUPAC Codes [4]). Three eumetazoan mitochondrial codes (TTs 2, 5 and 13) use only N, R and Y and no singlet codons, which causes their well-known high symmetry [23-25]. The non-standard codes extend the standard third position synonym doublets into triplets (see Figure 3, first eight rows of column 4) or reduce the standard third position synonym abN quartets into triplets; CUH leucine and UCB serine). All the reassignments, except the TT 22 (an algae) UCB serine, have either N, R, Y and H third position synonyms just like the standard code. The selection of third position synonymous codons from the 11 theoretical possibilities rather evolve to a further restriction of the selected ones (see TTs 2, 5 and 13) than to an adaptation to other third position synonyms such as for UCB serine of TT 22 (an algae). Seven reassignments don't form third position synonyms: In Figure 2, the three different reassignments of algae, the two Saccharomyces reassignments of TTs 12 and 26, and the TT 25 UGA glycine reassignment, and in Figure 1, the AGG lysine reassignment.

The strategy of the third position synonyms connects 61.1 % (11/18) of different reassignments to the standard code. The synonymous third position strategy covers the reassignments of the neighboring codon strategy from the previous Section 3.3.3.

Figure Number (TT)	Codon Reassignment	Neighboring Codons (Standard Codon Assignment)	Non-standard Synonyms
1 (14, 33)	UAA tyrosine	UAY (tyrosine)	UAH
2 (29)	UAR tyrosine	UAY (tyrosine)	UAN
1 (most) and 2 (5 TTs)	UGA tryptophan	UGG (tryptophan)	UGR
2 (10)	UGA cysteine	UGY (cysteine)	UGH
1 (2, 5, 13, 21) and 2 (3)	AUA methionine	AUG (methionine)	AUR
1 (9, 14, 21)	AAA asparagine	AAU (asparagine)	AAH
1 (5, 9, 14, 21)	AGR serine	AGU (serine)	AGN
1 (24, 33)	AGA serine	AGY (serine)	AGH
2 (6, 27, 28)	UAR glutamine	CAR (glutamine)	YAR
2 (30,31)	UAR glutamic acid	GAR (glutamic acid)	KAR
1 (13)	AGR glycine	GGN (glycine)	RGN (RGR, GGY)
1 (24, 33)	AGG lysine	AAR (lysine)	ARR (ARG, AAA)

Figure 3. Codon reassignments, neighboring codons standard and non-standard synonyms (Same color highlighting as in Figure 1 and Figure 2)

3.4. Reassignment Connections to the Standard Genetic Code

To what extent are the various strategies connected to the standard genetic code?

Most strategies connect to the standard code. Only the three algae reassignments have no specific connection to the standard code. In summary, 83.3 % (15/18) of the different reassignments have a specific connection to the standard code.

3.5. A Model for the Evolutionary Process Between the Vertebrate Mitochondrial and the Standard Code

3.5.1. Introduction

The easiest way to explain the undeniable eumetazoan TTs sequence, the eumetazoan unit structure and the unicellular reassignment connections to the eumetazoa is to use them as an evolutionary model. The various codon reassignments (Figure 1 and Figure 2) represent selected phyla and or classes of mostly heterotrophs, which make them a suitable frame for specific evolutionary research. Evolution conserves the standard genetic code from the prokaryotes on. Except for a few algae reassignments, autotrophs use the standard code also for their mitochondria. Heterotrophs deviate from the standard nuclear code of unicellulars for ciliates, flagellates and saccharomyces, and for selected unicellular and eumetazoan mitochondria. Most of the reassignments have an assisting function in the evolution from standard to vertebrate mitochondrial code. We summarize that function as sequence overlap (Section 3.1.1.) and unicellular-eumetazoan connecting reassignments (Section 3.2.2.). All the reassignment strategies contribute to the evolution of standard to vertebrate mitochondrial codes. The central and main sensitive codons for the investigation of this evolution are the AGG and UGA codons. Section 3.2.2. (the fourth connection) compares their reassignment sequences. UGA tryptophan, a vertebrate mitochondrial code property, originates in unicellular evolution from nuclear and mitochondrial UGA tryptophan reassignments. UGA tryptophan is also

present in all the eumetazoan mitochondria except in plants. A genetic code table (Figure 4) based on third position synonym codons shows the connections between the vertebrate mitochondrial genetic code and the standard genetic code (Figure 14 from [1]). References [1,26] describe this connection as a developmental transformation. Reference 1 contains interesting genetic code tables based on codon third position synonyms (Figures 14, 18 and 20 from [1]), which show special genetic code properties.

Vertebrate Mitochondrial Code				
An asterisk (*) indicates a difference for the standard code, which is shown in the lower part of the table.				
	abN encoded AAs		abY and abR (spliced abN) encoded AAs	
		ab	abY	abR
UC	Ser, S	UU	Phe, F	Leu, L
CU	Leu, L	UA	Tyr, Y	Stop
CC	Pro, P	UG	Cys, C	Trp, W*
CG	Arg, R	CA	His, H	Gln, Q
AC	Thr, T	AU	Ile, I*	Met, M*
GU	Val, V	AA	Asn, N	Lys, K
GC	Ala, A	AG	Ser, S	Stop*
GG	Gly, G	GA	Asp, D	Glu, E
No of AAs	8		7	8
Differences for the standard code (encoded AAs + stop)				
Third position nucleotides				
ab	G	A	H	R
UG	Trp, W	Stop		
AU	Met, M		Ile, I	
AG				Arg, R

“ab” is the first and the second (position) nucleotide of a codon triplet. U, C, A and G are the single letter abbreviations (IUPAC) for standard nucleotides. N, Y, R and H represent four, two pyrimidine, two purine and three (Y + A) third position nucleotides (IUPAC), respectively. AAs is amino acids. The 20 AAs are shown in three and one letter abbreviations. Turquoise blue and red highlights are for polar and non-polar amino acids (upper part of the figure), respectively. Pink highlight (only two numbers) is specific for the mitochondrial and green highlight is specific for the universal (standard) code. Amino acids that have six synonymous codons are in bold text.

Figure 4. Vertebrate mitochondrial and standard mRNA genetic code translations (Reproduced from American Journal of Educational Research 9 (1) 38-51, 2021 (reference [1]))

3.5.2. The Eumetazoan Unit Structure as an Evolutionary Model

The AGG reassignment sequence together with the assisting sequence overlap reassignments result in the TTs sequence of Figure 1, column 1. Our contribution proposes that the TTs sequence and all the eumetazoan reassignments of Figure 1 are the footprint of the eumetazoan evolutionary process from standard genetic code to the vertebrate mitochondrial code. We leave the details of the evolutionary interpretation of our results to the specialists. But it seems that the Platyhelminthes and the Echinodermata retain basic genetic properties of the common ancestor of the Bilateria clade. Bilateria have a mirror symmetrical left and right side at least during the embryonal development. From the Bilateria on, evolution diverges in the direction to the standard code by the lysine degeneracy compensation of the Hemichordata (Section 3.1.2.) and in the opposite way from the Platyhelminthes on to the vertebrate mitochondrial code. From the latter way on, the results are in line with evolutionary progress in the Bilateria clade from Acoelomates (Platyhelminthes), over Pseudocoelomates (Nematoda) and Coelomates (Arthropoda, Mollusca) to Tunicata, Chordata and Vertebrata.

4. Conclusions

4.1. The Non-Random Status of the Codon Reassignments

The codon reassignment strategies are intrinsic determinants for non-standard codes. They don't explain the reassignments in a mechanistic way by a gain or a loss of a tRNA or release factor. Reassignment strategies demonstrate a non-random status by showing clear and logical strategical connections between reassignments, their codons and/or the standard genetic code. Furthermore, the article confirms the non-random status by the fact that all TTs are interconnected firstly by the eumetazoan unit structure and secondly by a 100 % match between unicellular and eumetazoa reassignments. The 100 % match consists of a 76.9 % (Section 3.2.2., Comparable Bridging Sequences) and a 23.1 % contribution. The 23.1% contribution is discussed in Section 3.2.5., the Eumetazoan Connection and Contributions of The Termination Degeneracy Compensatory Strategy for Unicellular Algae. The restriction to 4 out of the 11 possibilities (and reduction to 3 for TTs 2, 5 and 13) for third position synonyms to N, Y, R (and H) make the strongly non-random connections of reassignments to TT 1 more evident. Not a single codon reassignment is random.

4.2. The Evolutionary Model

The reassignment connections to the standard genetic code and their non-random status are essential properties for the evolutionary process from the standard genetic code to the vertebrate mitochondrial code or vice versa.

The codon reassignments, especially the UGA and AGG bridging sequences, model the evolution for heterotrophs from the standard prokaryotes, reassigned

prokaryotes over specific ciliates / flagellates (eukaryote unicellular nuclear codes) following by mitochondria of protozoa and yeast to the mitochondrial eumetazoan reassignment generated TT sequence ending at the vertebrate mitochondrial code. This model forms a selected pathway through the heterotrophs.

In summary, the alternative genetic codes are the footprint of the evolution from standard to vertebrate mitochondrial code. Each codon reassignment is a step of this evolutionary process.

4.3. Significance

The most important significance of this contribution is the revelation of codon reassignment strategies, which prove the non-randomness of the alternative genetic codes. Furthermore, this investigation has evolutionary significance. Evolutionary specialists could connect the article's results to evolution theory in general and to the Margulis theory in particular [27].

Take Home Message

It is interesting to see how the results of molecular research can be put in a broader perspective. Sometimes strong beliefs can prevent one from finding such a broader connection.

Acknowledgments

The author would like to thank "The National Center for Biotechnology Information" (NCBI) for providing "The Genetic Codes" on-line and for the use that he could make of this data.

Abbreviations

TT(s) is translation table(s)

References

- [1] Struyf, J., "Parallelism between the Classical Geocentric Cosmos and the Life Chemistry Essentials" *American Journal of Educational Research*, 9, (1), 38-51, January 2021.
- [2] Nirenberg, M.W., "Deciphering the Genetic Code – a Personal Account" (Historical Review). *Trends in Biochemical Sciences* 29 (1). 46-54, January 2004.
- [3] Elzanowski, A., Ostell, J., "The Genetic Codes" Last update: Jan. 7 2019. <https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi>.
- [4] <https://www.bioinformatics.org/sms/iupac.html>.
- [5] Wichmann, S., Ardern, Z., "Optimality with standard genetic code is robust with respect to comparison code sets" *Biosystems* 185, 11, 104023, 2019.
- [6] Bezerra, A., Guimarães, A., Santos, M., "Non-Standard Genetic Codes Define New Concepts for Protein Engineering" *Life* 5 (4), 1610-1628, 2015.
- [7] Sammet, S., Bastolla, U., Porto, M., "Comparison of translation loads for standard and alternative genetic codes" *BMC Evol Biol* 10, 178, 2010.
- [8] Sengupta, S., Higgs, P., 2005 "A unified model of codon reassignment in alternative genetic codes" *Genetics* 170 (2): 831-840, 2005.
- [9] Link, J. A., Tirrell, D. A., "Reassignment of sense codons in vivo" *Methods*, 36 (3), 291-298, 2005.

- [10] Ring, K.L., Cavalcanti, A.R.O., Consequences of Stop Codon Reassignment on Protein Evolution in Ciliates with Alternative Genetic Codes, *Mol. Biol. and Evol.*, 25, (1), 179-186, 2008.
- [11] Silva, R.M., Miranda, I., Moura, G., Santos, M.A.S., Yeast as a model organism for studying the evolution of nonstandard genetic codes, Henry Stewart Publications 1473-9550. *Briefings In Functional Genomics And Proteomics*, 3 (1), 35-46, April 2004.
- [12] Tuite, M.F., Santos, M.A.S., Codon reassignment in *Candida* species: An evolutionary conundrum, *Biochimie* 78, 993-999, 1996.
- [13] Osawa, S., Jukes, T.H., Codon reassignment (codon capture) in evolution, *J Mol Evol.* 28 (4), 271-8, 1989.
- [14] Santos, M., Moura, G., Massey, S., Tuite, M., "Driving change: the evolution of alternative genetic codes" *Trends Genet* 20 (2): 95-102, 2004.
- [15] Osawa, S., Ohama, T., Jukes, T.H., Watanabe, K., Evolution of the mitochondrial genetic code. I. Origin of AGR serine and stop codons in metazoan mitochondria, *J Mol Evol.* 29 (3), 202-7, 1989.
- [16] Knight, R.D., Landweber, L.F., Yarus, M., How mitochondria redefine the code, *J Mol Evol.* 53 (4-5), 299-313, Oct-Nov 2001.
- [17] Andersson, S. G. E., Kurland, C. G., "An extreme codon preference strategy: codon reassignment". *Mol. Biol. Evol.*, 8 (4), 530-544, 1991.
- [18] Ma, N.J., Hemez, C.F., Barber, K.W., Rinehart, J., Isaacs, F.J., Organisms with alternative genetic codes resolve unassigned codons via mistranslation and ribosomal rescue, *eLife*. 7, 2018, <https://elifesciences.org/articles/34878>.
- [19] Sengupta, S., Xiaoguang Yang, X., Higgs, P.G., The Mechanisms of Codon Reassignments in Mitochondrial Genetic Codes, <https://arxiv.org/ftp/q-bio/papers/0703/0703066.pdf>.
- [20] Blazej, P., Wnętrzak, M., Mackiewicz, D., Gagat, P., Mackiewicz, P., Many alternative and theoretical genetic codes are more robust to amino acid replacements than the standard genetic code, *J. of Theoretical Biology*, 7, March 21-32, 2019.
- [21] Röttinger, E., and Lowe, C. J. "Evolutionary crossroads in developmental biology: hemichordates" *Development* 139: 2463-2475, 2012.
- [22] Lekomtsev, S. A., Nonstandard genetic codes and translation termination, *Molecular Biology* 41, 878-885, 2007.
- [23] Shu, J., 2017 "A new integrated symmetrical table for genetic codes" *BioSystems* 151: 21-26. 2017.
- [24] Shcherbak, V., "The co-operative symmetry of the genetic-code" *J. of Theoretical Biology* 132 (1): 121-124, 1988.
- [25] Findley, G., Findley, A., McGlynn, S., "Symmetry characteristics of the genetic code" *Proc Natl Acad Sci* 79 (22) 7061-7065, 1982.
- [26] Struyf, J., "The Human Hands Model for the Essentials of the Chemistry of Life" *World Journal of Chemical Education*, 6 (3): 117-123, 2018.
- [27] Margulis, L., *Origin of Eukaryotic Cells*, Yale University Press, 1970.



© The Author(s) 2021. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).