Journal of Theoretical Biology 464 (2019) 21–32

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/jtb

# Many alternative and theoretical genetic codes are more robust to amino acid replacements than the standard genetic code

Paweł Błażej, Małgorzata Wnętrzak, Dorota Mackiewicz, Przemysław Gagat, Paweł Mackiewicz*

*Department of Genomics, Faculty of Biotechnology, University of Wrocław, ul. Joliot-Curie 14a, 50-383, Wrocław, Poland*

ARTICLE INFO

*Article history:*
Received 1 October 2018
Revised 17 December 2018
Accepted 19 December 2018
Available online 21 December 2018

*Keywords:*
Alternative genetic codes
Amino acid
Mutation
Optimization
Standard genetic code

ABSTRACT

We evaluated the differences between the standard genetic code (SGC) and its known alternative variants in terms of the consequences of amino acids replacements. Furthermore, the properties of all the possible theoretical genetic codes, which differ from the SGC by one, two or three changes in codon assignments were also tested. Although the SGC is closer to the best theoretical codes than to the worst ones due to the minimization of amino acid replacements, from 10% to 27% of the all possible theoretical codes minimize the effect of these replacements better than the SGC. Interestingly, many types of codon reassignments observed in the alternative codes are also responsible for the substantial robustness to amino acid replacements. As many as 18 out of 21 alternatives perform better than the SGC under the assumed optimization criteria. These findings suggest that not all reassignments in the alternative codes are neutral and some of them could be selected to reduce harmful effects of mutations or translation of protein-coding sequences. The results also imply that the standard genetic code can be improved in this respect by a quite small number of changes, which are in fact realized in its variants. It would mean that the tendency to minimize mutational errors was not the main force that drove the evolution of the SGC.

© 2018 Elsevier Ltd. All rights reserved.

## 1. Introduction

The standard genetic code (SGC) has not been completely 'frozen' because many alternative variants have emerged from it. The alternative genetic codes operate mainly in mitochondrial genomes (Abascal et al., 2012; Boore and Brown, 1994; Clark-Walker and Weiller, 1994; Crozier and Crozier, 1993; Knight et al., 2001a; Osawa et al., 1989; Sengupta et al., 2007; Yokobori et al., 2003), but they are used by plastid (Del Cortona et al., 2017; Janouskovec et al., 2013; Lang-Unnasch and Aiello, 1999), nuclear (Hoffman et al., 1995; Panek et al., 2017; Sanchez-Silva et al., 2003; Santos et al., 1993; Schneider et al., 1989) and prokaryotic genomes as well, especially in parasites and symbionts (Bove, 1993; Campbell et al., 2013; Lim and Sears, 1992; McCutcheon et al., 2009). The alternative genetic codes seem to be particularly characteristic of small, multi-copy genomes that encode a limited number of proteins. Such genomes are perfect testing grounds for the genetic code evolution because codon reassignments may not be as harmful for them as for huge nuclear genomes, encoding thousands or tens of thousands of proteins (Massey and Garey, 2007).

The notable exception here are the nuclear genomes of ciliates, which despite their size and coding capacity did evolve the alternative genetic codes. They have, however, a peculiar genome architecture that greatly facilitates evolution, including deviations from the SGC (Bracht et al., 2013; Mochizuki, 2010). Their genetic information is contained in two types of nuclei: micronucleus and macronucleus, which serve as germline and somatic genomes, respectively, though the macronucleus can also transfer its acquired substitutions via a Lamarckian-type inheritance (Nowacki et al., 2008). The macronucleus is generated upon sexual reproduction from the micronucleus by extensive genome rearrangement and amplification. As a result, from ten standard chromosomes present in the micronucleus a pool of 20,000 multi-copy nanochromosomes are generated. They provide a perfect playground for evolutionary tinkering, especially taking into account amitotic division of the macronucleus during prevalent asexual reproduction (Bracht et al., 2013; Zufall et al., 2006).

In recent years, the number of newly discovered alternative genetic codes has significantly increased (Del Cortona et al., 2017; Heaphy et al., 2016; Muhlhausen et al., 2016; Zahonova et al., 2016). Until now, 31 alternative genetic codes have

* Corresponding author.
*E-mail address:* pamac@smorfland.uni.wroc.pl (P. Mackiewicz).

https://doi.org/10.1016/j.jtbi.2018.12.030
0022-5193/© 2018 Elsevier Ltd. All rights reserved.

been reported at the NCBI database (http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi), which is a good starting point to analyse their properties and evolutionary potential in comparison with the SGC.

All the deviations from the SGC that are observed in the alternative genetic codes can be divided into three main groups: (i) reassignments of codons encoding the typical 20 amino acids and stop translation signals, (ii) loss of codon meaning due to the disappearance of the codon itself and (iii) incorporation of new amino acids, e.g. selenocysteine and pyrrolysine. The majority of changes in the non-standard genetic codes belong to the first group and predominantly include reassignments of the stop (nonsense) codons to sense codons, e.g. UGA to tryptophan. The differences in the meaning of sense codons are less frequent and nearly always characteristic of mitochondrial genomes. Many of the same reassignments evolved independently in various phylogenetic lineages (Sengupta et al., 2007), and they can result from: (i) deletion or duplication of tRNA genes and associated mutations, e.g. in sites corresponding to anticodon and regions recognized by aminoacyl-tRNA synthetases, (ii) editing and posttranscriptional base modification of tRNA, (iii) mutations of genes for translational release factors and (iv) loss of codons due to strong mutational pressure (Knight et al., 2001b; Sengupta and Higgs, 2005; Sengupta et al., 2007).

The alternatives of the SGC are thought to have emerged via neutral evolution in small populations subjected to genetic drift and strong mutational pressure that led to tiny AT-rich genomes (Freeland et al., 2000; Sengupta and Higgs, 2015; Sengupta et al., 2007; Swire et al., 2005). On the other hand, it was proposed that codon changes associated with the deletion of tRNA genes are driven by selection to minimize the genome size and hence the time of replication (Andersson and Kurland, 1991; Andersson and Kurland, 1995). This genome streamlining hypothesis found support in a simulation model (Sengupta and Higgs, 2005) but not in the analyses of mitochondrial genomes (Knight et al., 2001a). The adaptive importance of codon reassignment was also proposed for AUA (Bender et al., 2008). In this case, the codon was recoded from isoleucine to methionine, and causes accumulation of methionine at the inner membrane of animal mitochondria that, in turn, plays anti-oxidant and cytoprotective role. The codon ambiguity can also promote phenotypic diversity and adaptability, e.g. it helps yeasts to cope with more stressful environments (Gomes et al., 2007; Santos et al., 1999). In turn, Swire et al. (2005) proposed that the mitochondrial genetic codes evolved to reduce protein synthesis costs by reassigning amino acids that are less expensive to synthesize.

In the light of the adaptive theory (Freeland and Hurst, 1998; Freeland et al., 2000; Haig and Hurst, 1991; Sonneborn, 1965; Woese, 1965), the standard genetic code evolved to minimize the consequences of point mutations or mistranslations, and it is very tempting to analyse the alternative codes in this context. However, the results of such studies have not been conclusive yet because their authors claimed that the alternative codes were not selected for mutational robustness and they are less adaptive than the SGC (Freeland et al., 2000; Kurnaz et al., 2010; Sammet et al., 2010) or they achieved a higher optimization level (Błażej et al., 2018; Morgens and Cavalcanti, 2013).

Since there is no consensus about the selection factors that drove the evolution of the alternative codes, we decided to investigate the properties of these codes in a more extensive approach, and compared them with the SGC according to error minimization in amino acid replacements. We created three groups of all possible theoretical genetic codes that differ from the standard genetic code by one, two or three codon reassignments. This large set of theoretical codes constituted a good reference for our comparisons. Thereby, we were able to test the consequences of codons reassignments and their impact on the robustness of genetic codes. Our results indicate that the codon reassignments observed in the non-standard genetic codes are generally responsible for improving their robustness in terms of amino acid replacements compared to the SGC. Moreover, many naturally occurring codon changes turned out to be close to the best theoretically possible ones.

## 2. Materials and methods

The properties of existing genetic codes were investigated using data available at the NCBI database (http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi). In total, we considered 21 non-standard genetic codes that differ from the SGC by at least one codon reassignment. We also wanted to study the genetic codes in which ambiguous codon reassignments are present. In order to do so, we considered two cases: all the changes or only those which are unambiguous. However, we excluded the Yeast Mitochondrial Code from the comparison of robustness of the genetic codes because this code has two codons unassigned in contrast to the other codes. The number of codon reassignments that occurred in evolution was collected from other papers (Knight et al., 2001b; Panek et al., 2017; Sengupta and Higgs, 2015; Sengupta et al., 2007) based on the most parsimonious phylogenetic distribution of these reassignments. The results were updated using newly discovered alternative genetic codes.

The robustness of a given code was measured by function $F$ that describes the consequences of amino acid and stop codon replacements introduced by a single point mutation. More precisely, the value of $F$, calculated for a selected genetic code, is expressed by the sum of squared differences in polarity values of encoded amino acids and is defined by the following formula:

$$F(code) = \sum_{<i,j>\in D} [f(i) - f(j)]^2, \tag{1}$$

where $D$ is the set of pairs of codons $i$ and $j$ that differ in one nucleotide substitution, whereas $f(i)$ and $f(j)$ are the polarity values of amino acids (Woese, 1973) encoded by the codons $i$ and $j$, respectively. Such a type of measure, with small modifications, has been used by many authors to estimate the optimality of the standard genetic code, e.g. Freeland and Hurst (1998), Haig and Hurst (1991) and Santos and Monteagudo (2010). In the formula, we also included single nucleotide substitutions that lead to nonsense mutations, i.e. to the replacement of an amino acid by a stop translation signal. In these cases we assumed that the effect of such a single replacement is equal to the maximum of squared differences calculated for any possible pair of amino acids.

For the genetic codes that contain ambiguous coding reassignments, we calculated three functions: (i) for the code assuming all the reassignments, both unambiguous and ambiguous $F(code\_all)$, (ii) for the code including only unambiguous reassignments $F(code\_unamb)$ and (iii) the average of these two values:

$$F(code\_avg) = \frac{1}{2}[F(code\_all) + F(code\_unamb)]. \tag{2}$$

We also compared the properties of non-standard genetic codes with three groups of theoretical genetic codes. The $G_1$, $G_2$ and $G_3$ group included all the possible codes that differ from the SGC by one, two or three codon reassignment, respectively. There are 1240 theoretical genetic codes in the $G_1$ group, 753,421 in $G_2$ and 299,066,360 in $G_3$. We considered several types of theoretical genetic codes regarding various restrictions placed on the stop codons: one, two or three stop codons can be recoded to amino acids as in some of the alternative genetic codes or all three stop codons are present as in the SGC. The analyses of all these theoretical possibilities enabled us to estimate the influence of both single and multiple codon reassignments on the robustness of the genetic code in the exhaustive approach.

Using the assumptions presented above, we made several comparisons between codes that differ in selected codon reassignments. For each group of theoretical genetic codes we computed basic descriptive statistics, such as: minimum, maximum and the median value of *F*. These parameters were compared with values obtained for the existing alternatives and the SGC. We also calculated the values of the parameter $B_i(test)$ for $i = 1, 2, 3$, which is the percentage of theoretical codes in each theoretical group that have the values of the function *F* smaller than the tested code, which could be the SGC, an alternative code or a theoretical code with a selected codon reassignment.

Moreover, we calculated the normalized percentage difference *Pd(test,SGC)* between the values of the function *F* for the SGC *F(SGC)* and the tested code *F(test)*, which could be an alternative code or a theoretical code with a selected reassignment. This difference is defined by:

$$Pd(test, SGC) = \frac{F(test) - F(SGC)}{F(SGC)} \cdot 100\%. \quad (3)$$

*Pd* can take values in the range from $-100$ to $+\infty$. In particular, negative values of *Pd* suggest that the SGC is worse, i.e. less robust in terms of the effects of amino acid replacements than the code *test*, whereas positive values of *Pd* indicate that the SGC minimizes the consequences better than the tested code.

We downloaded 4786 vertebrate mitochondrial genomes from the NCBI database (https://www.ncbi.nlm.nih.gov) and analysed in total 62,229 protein coding sequences annotated in these genomes. The nucleotide composition of four-fold degenerated sites (4FD) of these sequences was calculated using in-house written Perl scripts. The expected probability of occurrence of individual codons due to the mutational pressure was assessed by multiplication of appropriate nucleotide fractions in the 4FD sites. Spearman's correlation coefficient and log likelihood ratio (G-test) goodness of fit test with the conserved Yates correction were calculated in R package (R_Core_Team, 2017).

## 3. Results

### 3.1. The comparison of alternative genetic codes with the standard genetic code and its theoretical possibilities

The studied codes were compared according to the function *F*, which is the sum of squared differences in polarity values of encoded amino acids. Therefore, it can describe the consequences of amino acid replacements introduced by a point mutation. The smaller the value, the more robust the code is in terms of the effects of amino acid replacements. We also applied the percentage difference (*Pd*) between the values of the function *F* for the standard genetic code and the compared code. If *Pd* is negative it means that the SGC is less robust in terms of the consequences of amino acid replacements than the compared code. If it takes positive values, the SGC is better at minimization of the effects than the compared code.

In Table 1, we listed alternative genetic codes as well as the best and the worst theoretical codes in terms of the function *F* and *Pd*. Interestingly, for the majority of non-standard genetic codes, i.e. 18 out of 21, the effects of amino acid replacements are smaller than for the SGC. This bias significantly deviates from the equal distribution of the codes (G test, $p = 0.0014$). For the codes with ambiguous codon reassignments, smaller *F* values are observed when all codon reassignments are included (all), e.g. in such a case the *Condylostoma* and Karyorelict Nuclear Codes have the smallest *F* value of all the alternative codes (Table 1). These codes minimize the effects of amino acid replacements by 47.8% compared to the SGC. Sixteen alternative codes outperform the SGC by more than 10%. Only three non-standard genetic codes have worse robustness in terms of amino acid replacements than the SGC: the Vertebrate Mitochondrial Code, the Alternative Yeast Nuclear Code and the *Thraustochytrium* Mitochondrial Code. Just one codon change in the SGC, as it is in the case of the Candidate Division SR1 and the Gracilibacteria Code, can increase the robustness by 15.2%. Generally, the codes with more reassignments are characterized by better robustness than the SGC. The Spearman's correlation coefficient between the number of reassignments and *Pd* for the whole data set is negative and equals $-0.48$, and it is also statistically significant ($p = 0.013$). The relationship is more pronounced when the SGC and the codes better than it, i.e. with $F \leq 5641.46$, are analysed: $\rho = -0.55$, $p = 0.006$.

The approach presented above was focused on analysing the alternative codes with reference to the standard genetic code. It is also noteworthy to compare these codes with the theoretical genetic codes that differ from the SGC by a fixed number of codon

**Table 1**
The standard, alternative and theoretical genetic codes arranged according to their values of the function *F*.

| Genetic code | *N* | *F* | *Pd* |
|---|---|---|---|
| BTC with no stop codons | 3 | 2861.46 | −49.28 |
| *Condylostoma* Nuclear (all) | 3 | 2945.12 | −47.80 |
| Karyorelict Nuclear (all) | 3 | 2945.12 | −47.80 |
| Karyorelict Nuclear (mean) | 3 | 3484 | −38.24 |
| BTC with at least one stop codon | 3 | 3677.16 | −34.82 |
| *Blastocrithidia* Nuclear (all) | 3 | 3697.04 | −34.47 |
| BTC with at least one stop codon | 2 | 3940.02 | −30.16 |
| Alternative Flatworm Mitochondrial | 5 | 3942.66 | −30.11 |
| CDH Nuclear | 2 | 4022.88 | −28.69 |
| Karyorelict Nuclear (unamb) | 2 | 4022.88 | −28.69 |
| *Mesodinium* Nuclear | 2 | 4116.26 | −27.04 |
| *Blastocrithidia* Nuclear (mean) | 3 | 4250.7 | −24.65 |
| *Condylostoma* Nuclear (mean) | 3 | 4293.29 | −23.90 |
| WTC with no stop codons | 3 | 4326.56 | −23.31 |
| Peritrich Nuclear | 2 | 4692.9 | −16.81 |
| Trematode Mitochondrial | 5 | 4692.92 | −16.81 |
| Invertebrate Mitochondrial | 4 | 4696.84 | −16.74 |
| Ascidian Mitochondrial | 4 | 4704.2 | −16.61 |
| Echinoderm and Flatworm Mitochondrial | 4 | 4706.84 | −16.57 |
| BTC with at least one stop codon | 1 | 4766.28 | −15.51 |
| Candidate Division SR1 and Gracilibacteria | 1 | 4783.56 | −15.21 |
| Euplotid Nuclear | 1 | 4795.08 | −15.00 |
| *Blastocrithidia* Nuclear (unamb) | 1 | 4804.36 | −14.84 |
| MPC Mitochondrial and M/S | 1 | 4804.36 | −14.84 |
| Pterobranchia Mitochondrial | 3 | 4839.88 | −14.21 |
| Chlorophycean Mitochondrial | 1 | 4936.3 | −12.50 |
| BTC with fixed stop codons | 3 | 5116.3 | −9.31 |
| BTC with fixed stop codons | 2 | 5206.34 | −7.71 |
| BTC with fixed stop codons | 1 | 5378.6 | −4.66 |
| *Scenedesmus obliquus* Mitochondrial | 2 | 5575.14 | −1.18 |
| *Pachysolen tannophilus* Nuclear | 1 | 5630.96 | −0.19 |
| **Standard** | **0** | **5641.46** | **0.00** |
| Alternative Yeast Nuclear | 1 | 5651.86 | 0.18 |
| *Thraustochytrium* Mitochondrial | 1 | 6283.02 | 11.37 |
| WTC with fixed stop codons | 1 | 6668.66 | 18.21 |
| Vertebrate Mitochondrial | 4 | 6716.48 | 19.06 |
| WTC with at least one stop codon | 1 | 6807.66 | 20.67 |
| WTC with fixed stop codons | 2 | 7519.16 | 33.28 |
| WTC with at least one stop codon | 2 | 7961.94 | 41.13 |
| WTC with fixed stop codons | 3 | 8353.46 | 48.07 |
| WTC with at least one stop codon | 3 | 9112.2 | 61.52 |

*N* is the number of reassignments; *Pd* is the normalized percentage difference of the alternative and theoretical codes compared with the standard genetic code (SGC). In the case of the alternative genetic codes containing ambiguous coding reassignments, we calculated the parameters for: (i) the code assuming all reassignments, both unambiguous and ambiguous (all) and (ii) the code with only unambiguous reassignments (unamb). We also averaged these two values (mean). The *Condylostoma* Nuclear Code differs from the SGC only in ambiguous reassignments. Therefore, its unambiguous version is the same as the SGC. BTC - the best theoretical code; WTC - the worst theoretical code; CDH - Ciliate, Dasycladacean and *Hexamita*; MPC Mitochondrial and M/S - Mold, Protozoan, and Coelenterate Mitochondrial Code and the *Mycoplasma*/*Spiroplasma* Code.

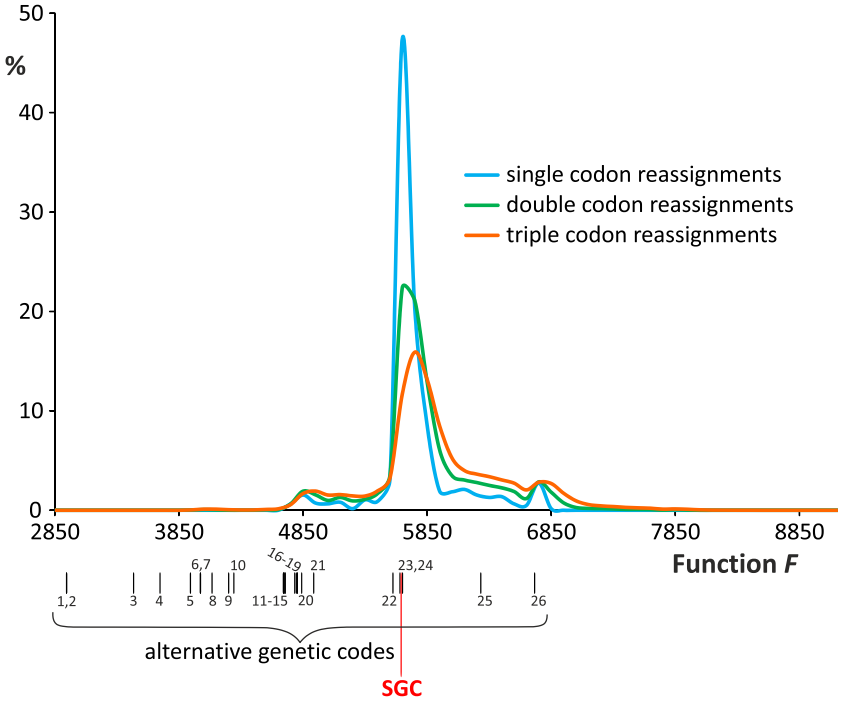


**Fig. 1.** The distribution of the function *F* for three types of theoretical codes. All theoretical codes differing by one, two and three codon reassignments from the standard genetic code (SGC) were considered. The *F* values for the standard genetic code (SGC) and alternative genetic codes were marked by lines under the x axis: 1 - *Condylostoma* Nuclear (all); 2 - Karyorelict Nuclear (all); 3 - Karyorelict Nuclear (mean); 4 - *Blastocrithidia* Nuclear (all); 5 - Alternative Flatworm Mitochondrial; 6 - Ciliate, Dasycladacean and *Hexamita* Nuclear; 7 - Karyorelict Nuclear (unamb); 8 - *Mesodinium* Nuclear; 9 - *Blastocrithidia* Nuclear (mean); 10 - *Condylostoma* Nuclear (mean); 11 - Peritrich Nuclear; 12 - Trematode Mitochondrial; 13 - Invertebrate Mitochondrial; 14 - Ascidian Mitochondrial; 15 - Echinoderm and Flatworm Mitochondrial; 16 - Candidate Division SR1 and Gracilibacteria; 17 - Euplotid Nuclear; 18 - *Blastocrithidia* Nuclear (unamb); 19 - Mold, Protozoan, and Coelenterate Mitochondrial Code and the *Mycoplasma*/*Spiroplasma*; 20 - Pterobranchia Mitochondrial; 21 - Chlorophycean Mitochondrial; 22 - *Scenedesmus obliquus* Mitochondrial; 23 - *Pachysolen tannophilus* Nuclear; 24 - Alternative Yeast Nuclear; 25 - *Thraustochytrium* Mitochondrial; 26 - Vertebrate Mitochondrial. The *F* function is a measure of the code robustness to amino acid replacements; the smaller the value, the better the code minimizes the consequences of these replacements. It is clear that the majority of alternative codes are characterized by smaller *F* values than the SGC and most theoretical codes.

changes. It allows to place the real codes in the global space of possible codes. Therefore, we took into consideration three groups of codes differing by one, two or three codon reassignments from the SGC. Moreover, three other types of theoretical genetic codes were studied. The first one has all three stop codons fixed as in the SGC, the second one maintains at least one unrestricted stop codon and the last one does not have any stop codons. Table 1 includes the results for the codes that obtained the smallest values of the function *F* (the best theoretical codes, BTC) and the greatest values of the function *F* (the worst theoretical codes, WTC), while Fig. 1 shows the distribution of the *F* values for all the theoretical codes with markings of the *F* values for the alternative genetic codes and the SGC.

The distribution of the function *F* for the theoretical codes is characterized by a relatively narrow high peak with quite flat and long tails (Fig. 1). The distribution is not perfectly symmetric because there are slightly more codes with greater values of *F* to the right of the peak. The value for the SGC, i.e. 5641 is only slightly below the medians for the theoretical codes with a single, double and triple codon reassignments, i.e. 5681, 5760 and 5845, respectively (Fig. 2). The SGC value is in the first and the third quartile range (5638–5785) of the theoretical codes with a single codon reassignment but a bit below the quartile ranges 5657–5970 and 5683–6216 of the theoretical codes with double and triple codon reassignments, respectively. The SGC is closer to the best theoretical solutions than to the worst theoretical codes but the distance to the minimum grows with the number of codon reassignments in the theoretical codes (Fig. 2). On the other hand, the majority of alternative codes have the values of the function *F* much smaller than the median and out of the quartile ranges of the theoretical

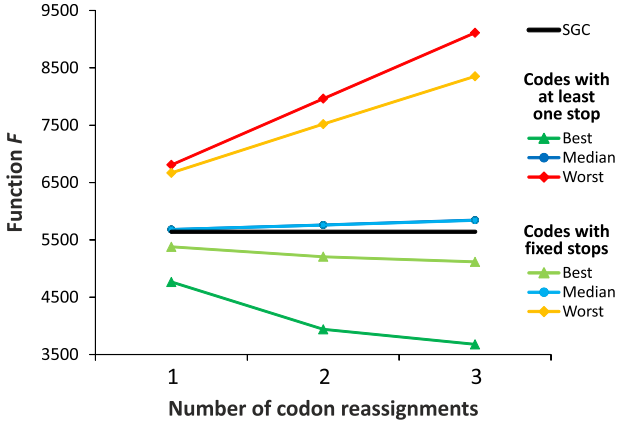


**Fig. 2.** The values of the function *F* for theoretical genetic codes and the standard genetic code (SGC). The theoretical genetic codes with at least one stop codon and with the stop codons fixed as in the SGC were considered. The values for the best and the worst codes as well as the median were shown.

codes (Fig. 1). The non-standard codes are also much closer to the theoretical minimum than to the SGC.

The best codes with a single codon reassignment are slightly better than the SGC by 4.7% to 9.3% but the codes with a double codon reassignment have the values of the function *F* substantially smaller than the SGC and outperform it by 15.5% to 34.8% (Table 1). In turn, the SGC is better than the worst theoretical codes by 18.2% to 61.5%. As expected, the biggest impact on the function *F* is exerted by the stop codons. Therefore, the theoretical codes with

**Table 2**
The percentage of theoretical codes ($B_i$) having the $F$ values smaller than the standard genetic code.

| Genetic code | $F$ | $B_1$ | $B_2$ | $B_3$ |
|---|---|---|---|---|
| *Condylostoma* Nuclear (all) | 2945.12 | 0 | 0 | 0 |
| Karyorelict Nuclear (all) | 2945.12 | 0 | 0 | 0 |
| Karyorelict Nuclear (mean) | 3484 | 0 | 0 | 0 |
| *Blastocrithidia* Nuclear (all) | 3697.04 | 0 | 0 | 5.3E-05 |
| Alternative Flatworm Mitochondrial | 3942.66 | 0 | 8.0E-04 | 0.01 |
| CDH Nuclear | 4022.88 | 0 | 0.04 | 0.06 |
| Karyorelict Nuclear (unamb) | 4022.88 | 0 | 0.04 | 0.06 |
| *Mesodinium* Nuclear | 4116.26 | 0 | 0.09 | 0.17 |
| *Blastocrithidia* Nuclear (mean) | 4250.7 | 0 | 0.13 | 0.29 |
| *Condylostoma* Nuclear (mean) | 4293.29 | 0 | 0.13 | 0.31 |
| Peritrich Nuclear | 4692.9 | 0 | 0.25 | 0.58 |
| Trematode Mitochondrial | 4692.92 | 0 | 0.25 | 0.58 |
| Invertebrate Mitochondrial | 4696.84 | 0 | 0.26 | 0.59 |
| Ascidian Mitochondrial | 4704.2 | 0 | 0.27 | 0.61 |
| Echinoderm and Flatworm Mitochondrial | 4706.84 | 0 | 0.27 | 0.61 |
| Candidate Division SR1 and Gracilibacteria | 4783.56 | 0.32 | 0.68 | 0.99 |
| Euplotid Nuclear | 4795.08 | 0.56 | 0.83 | 1.11 |
| *Blastocrithidia* Nuclear (unamb) | 4804.36 | 0.64 | 0.99 | 1.22 |
| MPC Mitochondrial and M/S | 4804.36 | 0.64 | 0.99 | 1.22 |
| Pterobranchia Mitochondrial | 4839.88 | 1.69 | 1.71 | 1.75 |
| Chlorophycean Mitochondrial | 4936.3 | 2.74 | 3.54 | 3.58 |
| *Scenedesmus obliquus* Mitochondrial | 5575.14 | 7.41 | 12.25 | 14.68 |
| *Pachysolen tannophilus* Nuclear | 5630.96 | 19.90 | 18.60 | 18.34 |
| **Standard** | **5641.46** | **27.40** | **21.11** | **19.45** |
| Alternative Yeast Nuclear | 5651.86 | 34.81 | 23.64 | 21.57 |
| *Thraustochytrium* Mitochondrial | 6283.02 | 93.23 | 85.21 | 77.24 |
| Vertebrate Mitochondrial | 6716.48 | 97.26 | 93.78 | 89.39 |

The theoretical codes with at least one stop codon and characterized by one ($B_1$), two ($B_2$) or three ($B_3$) different codon reassignments in comparison with the standard genetic code were considered. CDH - Ciliate, Dasycladacean and *Hexamita*; MPC Mitochondrial and M/S - Mold, Protozoan, and Coelenterate Mitochondrial Code and the *Mycoplasma*/*Spiroplasma* Code.

triple codon reassignments and no stop translation signal encoded are characterized by the smallest values of the $F$ function and the largest $Pd$ (49.3%). Such codes outperform the SGC by 23.3% even if we take into account their worst case (Table 1). The alternative codes compare quite favourably with the theoretical ones. Many of the alternatives have their robustness close to that of the best theoretical codes even if we take into account the codes with the same number of reassignments.

In order to compare quantitatively the standard and alternative genetic codes with the theoretical possibilities, we calculated the percent of appropriate theoretical codes with better robustness in terms of amino acid replacements than the codes in question. There are only 9.8% of theoretical codes without stop codons better than the *Condylostoma* and Karyorelict Nuclear Codes assuming recoding of all three stop signals to amino acids. However, for another code of this type, *Blastocrithidia* Nuclear Code, as much as 95.4% of theoretical codes have smaller values of the $F$ function than this alternative code.

Other alternative codes are more suitable to compare with the theoretical codes in which at least one codon encodes the stop translation signal (Table 2). Surprisingly, there are many alternative codes for which there is a very small percentage of the theoretical codes with smaller value of the function $F$. For 13 alternative codes with at least one stop codon, this percentage is below 4%, and for 9 codes it is below 1%, even if we take into account the theoretical codes which differ from the SGC by three codon changes. However, there are also two alternative genetic codes which are worse than over 90% of theoretical alternatives.

In terms of the robustness, the standard genetic code is surpassed by 25% (280), 15% (92,476) and 10.4% (22,815,206) of the theoretical codes having the same fixed stop codons and one, two or three missense changes, respectively. The codes with better robustness than the SGC and with at least one stop codon are more numerous, i.e. 27.4% (340), 21.1% (159,040) and 19.5% (58,313,518).

### 3.2. Properties of codon reassignments in terms of code robustness

Most reported reassignment types, i.e. 13 out of 27, are missense ones, which change the coding of one amino acid to another. There are 10 reassignments recoding a stop codon to a sense codon and four reassignments changing a sense codon to a stop codon. However, these reassignment types are not equally distributed in the alternative codes. Many of them evolved independently in various phylogenetic lineages. The most frequent is the reassignment of the stop codon TGA to tryptophan, which evolved in at least 17 codes.

There is a significant predominance of single reassignments that can improve the SGC in term of the robustness (Table 3), i.e. 61 versus 6 (G test, $p = 1.4e{-}12$). Forty-four of them can increase the SGC robustness by more than 5%. The biased distribution concerns also the missense reassignments. There are 15 such beneficial codon changes versus only 2 deleterious ones (G test, $p = 0.002$). The changes that make the code more robust evolve more often than the changes deteriorating its property. Accordingly, the Spearman's correlation coefficient between the number of single reassignments and the $F$ values is negative $-0.53$ and statistically significant ($p = 0.004$).

The codon changes in the alternative codes turn out quite well also in comparison with all the theoretical codes assuming at least one stop codon and differing from the SGC by one codon reassignment (Table 3). Sixty-one out of all 67 reassignments are in the first quarter of the best theoretical reassignments and for 48 reassignments there are no more than 10% of better theoretical possibilities. The recoding of the stop codon TGA to alanine is responsible for the biggest drop in the value of the function $F$ but it was not observed in the alternative codes. However, the codes with the top best real reassignments can also improve the SGC to a similar extent, i.e. by about 15%. These reassignments involve recoding stop codons to sense codons but the real missense changes,

**Table 3**
The single codon reassignments observed in alternative genetic codes.

| Reassignment | Type | *R* | *F* | *Pd* | *B* |
|---|---|---|---|---|---|
| TGA(Stp)→Ala | min nonsense | 0 | 4766.28 | −15.51 | 0 |
| TGA(Stp)→Gly | alternative | 2 | 4783.56 | −15.21 | 0.32 |
| TGA(Stp)→Cys | alternative | 1 | 4795.08 | −15.00 | 0.56 |
| TGA(Stp)→Trp | alternative | 17 | 4804.36 | −14.84 | 0.65 |
| TAG(Stp)→Gln | alternative | 8 | 4843.06 | −14.15 | 1.69 |
| TAG(Stp)→Tyr | alternative | 1 | 4879.02 | −13.51 | 1.94 |
| TAG(Stp)→Leu | alternative | 3 | 4936.3 | −12.50 | 2.74 |
| TAA(Stp)→Gln | alternative | 7 | 5083.72 | −9.89 | 3.15 |
| TAA(Stp)→Tyr | alternative | 3 | 5141.14 | −8.87 | 3.63 |
| TAG(Stp)→Glu | alternative | 2 | 5219.02 | −7.49 | 4.35 |
| TAA(Stp)→Glu | alternative | 2 | 5377.78 | −4.67 | 4.68 |
| GAT(Asp)→Arg | min missense | 0 | 5378.6 | −4.66 | 4.84 |
| AGG(Arg)→Ser | alternative | 1 | 5601.14 | −0.71 | 9.68 |
| AGG(Arg)→Gly | alternative | 1 | 5602.58 | −0.69 | 9.92 |
| AGA(Arg)→Gly | alternative | 1 | 5616.02 | −0.45 | 11.69 |
| AGA(Arg)→Ser | alternative | 1 | 5617.78 | −0.42 | 12.10 |
| CTG(Leu)→Thr | alternative | 1 | 5620.72 | −0.37 | 14.03 |
| CTA(Leu)→Thr | alternative | 1 | 5623.44 | −0.32 | 15.89 |
| CTC(Leu)→Thr | alternative | 1 | 5624.12 | −0.31 | 16.45 |
| CTT(Leu)→Thr | alternative | 1 | 5624.12 | −0.31 | 16.45 |
| ATA(Ile)→Met | alternative | 4 | 5624.82 | −0.29 | 16.85 |
| CTG(Leu)→Ala | alternative | 1 | 5630.96 | −0.19 | 19.92 |
| AAA(Lys)→Asn | alternative | 2 | 5638.02 | −0.06 | 23.95 |
| CTG(Leu)→Ser | alternative | 1 | 5651.86 | 0.18 | 34.76 |
| AGG(Arg)→Lys | alternative | 1 | 5713.46 | 1.28 | 60.48 |
| TCA(Ser)→Stp | alternative | 1 | 6280.3 | 11.32 | 93.06 |
| TTA(Leu)→Stp | alternative | 1 | 6283.02 | 11.37 | 93.23 |
| AGA(Arg)→Stp | alternative | 1 | 6497.1 | 15.17 | 95.81 |
| TTT(Phe)→Asp | max missense | 0 | 6668.66 | 18.21 | 96.77 |
| AGG(Arg)→Stp | alternative | 1 | 6737.06 | 19.42 | 97.50 |
| TTT(Phe)→Stp | max nonsense | 0 | 6807.66 | 20.67 | 100 |

The single codon reassignments were compared with the selected reassignments responsible for the biggest drop (min) and the biggest increase (max) in the value of the function *F*. *R* is the number of independently evolved reassignments in the alternative codes. *F* is the *F* function value calculated for the genetic codes differing from the standard genetic code (SGC) in the given reassignment. *Pd* means the percentage difference of these codes from the SGC. *B* is the percent of theoretical codes assuming at least one stop codon and having the *F* values smaller than the code with the given reassignment.

**Table 4**
The single codon reassignments recoding a stop codon to an amino acid in alternative genetic codes.

| Reassignment | Type | *R* | *F* | *Pd* | *B* |
|---|---|---|---|---|---|
| TGA(Stp)→Ala | min | 0 | 4766.28 | −15.51 | 0.00 |
| TGA(Stp)→Gly | alternative | 2 | 4783.56 | −15.21 | 6.67 |
| TGA(Stp)→Cys | alternative | 1 | 4795.08 | −15.00 | 11.67 |
| TGA(Stp)→Trp | alternative | 17 | 4804.36 | −14.84 | 13.33 |
| TAG(Stp)→Gln | alternative | 8 | 4843.06 | −14.15 | 35.00 |
| TAG(Stp)→Tyr | alternative | 1 | 4879.02 | −13.51 | 40.00 |
| TAG(Stp)→Leu | alternative | 3 | 4936.3 | −12.50 | 56.67 |
| TAA(Stp)→Gln | alternative | 7 | 5083.72 | −9.89 | 65.00 |
| TAA(Stp)→Tyr | alternative | 3 | 5141.14 | −8.87 | 75.00 |
| TAG(Stp)→Glu | alternative | 2 | 5219.02 | −7.49 | 90.00 |
| TAA(Stp)→Glu | alternative | 2 | 5377.78 | −4.67 | 96.67 |
| TAA(Stp)→Asp | max | 0 | 5446.28 | −3.46 | 100 |

The single codon reassignments were compared with the selected reassignments responsible for the biggest drop (min) and the biggest increase (max) in the value of the function *F*. *B* is the percentage of theoretical codes recoding one stop codon to an amino acid and having the value of the function *F* smaller than the code with the given reassignment. Other columns as in Table 3.

i.e. excluding the stop codons, are also quite profitable because 15 out of 17 such changes are in the first quartile of the best theoretical reassignments.

Since the reassignments involving the stop codons are special ones, we also compared them with other changes of the same type, i.e. recoding one stop codon to an amino acid (Table 4). Many of the real reassignments are still better than other possibilities, e.g. 29 out of 46 are in the best 40% of the theoretical reassignments.

Since the alternative genetic codes have usually more than one reassignment, we also analysed double and triple combinations of the single codon changes present in a given code. Moreover, to assess the quality of these reassignments among the potential possibilities, we studied the theoretical genetic codes that differ from the SGC in each double and triple combination of the single codon reassignments.

In the case of double codon changes (Table 5), there also exists a significant number of codes, i.e. 62 versus 6, that have negative percentage difference compared to the SGC (G test, $p = 7.6e{-}13$). Thirty-four of them can improve the SGC by more than 10%. The missense reassignments also show this trend because almost all of them, i.e. 20 versus 1, have $Pd < 0$ (G test, $p = 1.\ 9e{-}5$). All the changes turned out quite beneficial in comparison with the theoretical codes that assumed two reassignments. Forty out of 68 codon changes in the alternative codes are in the first 6% of the theoretical reassignments generating the smallest values of the function *F*, while almost half, i.e. 33 reassignments, are in the first 1% of the best theoretical changes.

The best theoretical reassignment recodes two stop codons to amino acids and can lead to 30% improvement of the SGC (Table 5). Similar values characterize the best reassignments in the alternative codes. The normalized percentage distance for the extreme theoretical reassignments, excluding stop codons and stop signals, is −7.7% and 33% for the best and the worst reassignment, respectively. The observed changes in the alternative codes are much closer to the former rather than the latter value and are in the range from −1.32% to 0.97%. The real reassignments recoding two stop codons to sense codons turned out also quite well because 12 out of 19 reassignments are in the first quartile of the best theoretical possibilities when compared with the same type of theoretical changes (Table 6).

Most combinations of three reassignments occurring in the alternative genetic codes also minimize the function *F* better than the SGC (Table 7). There are 40 such combinations versus 4, which significantly deviate from the equal distribution (G test, $p = 1.6e{-}8$). The same is also true for the missense reassignments because all such changes have negative *Pd* values (G test, $p = 0.014$). Forty out of 54 codes with the triple reassignments are in the best 7% of the theoretical codes.

The best theoretical improvement of the SGC robustness using three reassignments gave us the percentage difference 49.3%, which is very similar to −47.8%, i.e. the *Pd* value found for *Condylostoma* and Karyorelict Nuclear codes assuming that all three stop codons can code for amino acids (Table 7). These code versions are also in the first 10% of the best possible reassignments removing three stop codons from the code (Table 8). The best and the worst theoretical codon reassignments involving only sense codons and amino acids would cause 9.3% improvement and 48.1% deterioration of the SGC, respectively. However, all the reassignments of this type observed in the natural alternative codes are only beneficial and can improve the SGC by 1.6% to 0.74%.

## 4. Discussion

Previous analyses on the robustness of the alternative genetic codes concluded that they are generally not better optimized regarding point mutations than the standard genetic code, and they evolved rather under neutral evolution without selection on the optimality (Freeland et al., 2000; Kurnaz et al., 2010; Sammet et al., 2010). Depending on the assumed transition to transversion ratio (w), Freeland et al. (2000) found that each of the 13 considered alternative codes is slightly worse than the SGC (for $w = 1$) and six of them are as good as the SGC (for $w = 5$). They

**Table 5**
The double codon reassignments observed in alternative genetic codes.

| Reassignment | Type | *R* | *F* | *Pd* | *B* |
|---|---|---|---|---|---|
| TAA(Stp)→Gly; TGA(Stp)→Ala | min nonsense | 0 | 3940.02 | −30.16 | 0 |
| TAA(Stp)→Gln; TGA(Stp)→Trp | alternative | 2 | 4005.96 | −28.99 | 0.02 |
| TAG(Stp)→Gln; TGA(Stp)→Trp | alternative | 2 | 4005.96 | −28.99 | 0.02 |
| TAA(Stp)→Gln; TAG(Stp)→Gln | alternative | 8 | 4022.88 | −28.69 | 0.04 |
| TAA(Stp)→Tyr; TGA(Stp)→Trp | alternative | 2 | 4041.92 | −28.35 | 0.05 |
| TAA(Stp)→Tyr; TAG(Stp)→Tyr | alternative | 1 | 4116.26 | −27.04 | 0.09 |
| TAA(Stp)→Glu; TGA(Stp)→Trp | alternative | 1 | 4381.92 | −22.33 | 0.13 |
| TAG(Stp)→Glu; TGA(Stp)→Trp | alternative | 1 | 4381.92 | −22.33 | 0.13 |
| TAA(Stp)→Glu; TAG(Stp)→Glu | alternative | 2 | 4692.9 | −16.81 | 0.25 |
| AGA(Arg)→Ser; TGA(Stp)→Trp | alternative | 1 | 4761.48 | −15.6 | 0.43 |
| AGA(Arg)→Gly; TGA(Stp)→Trp | alternative | 1 | 4763.56 | −15.56 | 0.46 |
| AGG(Arg)→Ser; TGA(Stp)→Trp | alternative | 1 | 4764.04 | −15.55 | 0.46 |
| AGG(Arg)→Gly; TGA(Stp)→Trp | alternative | 1 | 4765.48 | −15.53 | 0.48 |
| CTG(Leu)→Thr; TGA(Stp)→Trp | alternative | 1 | 4783.62 | −15.21 | 0.68 |
| CTA(Leu)→Thr; TGA(Stp)→Trp | alternative | 1 | 4786.34 | −15.16 | 0.71 |
| CTC(Leu)→Thr; TGA(Stp)→Trp | alternative | 1 | 4787.02 | −15.15 | 0.72 |
| CTT(Leu)→Thr; TGA(Stp)→Trp | alternative | 1 | 4787.02 | −15.15 | 0.72 |
| ATA(Ile)→Met; TGA(Stp)→Trp | alternative | 4 | 4787.72 | −15.13 | 0.74 |
| AAA(Lys)→Asn; TGA(Stp)→Trp | alternative | 2 | 4800.92 | −14.9 | 0.92 |
| AGG(Arg)→Lys; TGA(Stp)→Trp | alternative | 1 | 4876.36 | −13.56 | 2.39 |
| AGG(Arg)→Ser; TAA(Stp)→Tyr | alternative | 2 | 5100.82 | −9.58 | 5.37 |
| AGA(Arg)→Ser; TAA(Stp)→Tyr | alternative | 2 | 5117.46 | −9.29 | 5.55 |
| TAA(Stp)→Tyr; ATA(Ile)→Met | alternative | 1 | 5124.5 | −9.16 | 5.63 |
| AAA(Lys)→Asn; TAA(Stp)→Tyr | alternative | 1 | 5135.96 | −8.96 | 5.78 |
| GAA(Glu)→Arg; GAT(Asp)→Gln | min missense | 0 | 5206.34 | −7.71 | 6.73 |
| AGA(Arg)→Ser; AGG(Arg)→Ser | alternative | 1 | 5567.22 | −1.32 | 11.95 |
| AGA(Arg)→Gly; AGG(Arg)→Gly | alternative | 1 | 5571.38 | −1.24 | 12.09 |
| TAG(Stp)→Leu; TCA(Ser)→Stp | alternative | 1 | 5575.14 | −1.18 | 12.25 |
| ATA(Ile)→Met; AGG(Arg)→Ser | alternative | 2 | 5584.5 | −1.01 | 12.75 |
| ATA(Ile)→Met; AGG(Arg)→Gly | alternative | 1 | 5585.94 | −0.98 | 12.82 |
| AAA(Lys)→Asn; AGG(Arg)→Ser | alternative | 2 | 5597.7 | −0.78 | 13.63 |
| ATA(Ile)→Met; AGA(Arg)→Gly | alternative | 1 | 5601.3 | −0.71 | 13.98 |
| ATA(Ile)→Met; AGA(Arg)→Ser | alternative | 2 | 5603.7 | −0.67 | 14.23 |
| CTG(Leu)→Thr; ATA(Ile)→Met | alternative | 2 | 5604.08 | −0.66 | 14.27 |
| CTA(Leu)→Thr; ATA(Ile)→Met | alternative | 2 | 5604.08 | −0.66 | 14.27 |
| CTC(Leu)→Thr; ATA(Ile)→Met | alternative | 2 | 5607.48 | −0.60 | 14.67 |
| CTT(Leu)→Thr; ATA(Ile)→Met | alternative | 2 | 5607.48 | −0.60 | 14.67 |
| AAA(Lys)→Asn; AGA(Arg)→Ser | alternative | 2 | 5613.7 | −0.49 | 15.42 |
| AAA(Lys)→Asn; ATA(Ile)→Met | alternative | 1 | 5621.54 | −0.35 | 16.78 |
| AGA(Arg)→Ser; AGG(Arg)→Lys | alternative | 1 | 5696.18 | 0.97 | 35.26 |
| AGA(Arg)→Stp; TGA(Stp)→Trp | alternative | 1 | 5893.56 | 4.47 | 69.69 |
| AGG(Arg)→Stp; TGA(Stp)→Trp | alternative | 1 | 5899.96 | 4.58 | 70.27 |
| ATA(Ile)→Met; AGA(Arg)→Stp | alternative | 1 | 6486.86 | 14.99 | 90.08 |
| ATA(Ile)→Met; AGG(Arg)→Stp | alternative | 1 | 6720.42 | 19.13 | 93.85 |
| AGA(Arg)→Stp; AGG(Arg)→Stp | alternative | 1 | 7330.26 | 29.94 | 99.57 |
| TTT(Phe)→Asp; CTC(Leu)→Asp | max missense | 0 | 7519.16 | 33.28 | 99.77 |
| TTT(Phe)→Stp; CCT(Pro)→Stp | max nonsense | 0 | 7961.94 | 41.13 | 100 |

The double codon reassignments were compared with the selected reassignments responsible for the biggest drop (min) and the biggest increase (max) in the values of the function *F*. Other columns as in Table 3.

**Table 6**
The double codon reassignments recoding a stop codon to an amino acid in alternative genetic codes.

| Reassignment | Type | *R* | *F* | *Pd* | *B* |
|---|---|---|---|---|---|
| TAA(Stp)→Gly; TGA(Stp)→Ala | Min | 0 | 3940.02 | −30.16 | 0.00 |
| TAA(Stp)→Gln; TGA(Stp)→Trp | Alternative | 2 | 4005.96 | −28.99 | 15.17 |
| TAG(Stp)→Gln; TGA(Stp)→Trp | Alternative | 2 | 4005.96 | −28.99 | 15.17 |
| TAA(Stp)→Gln; TAG(Stp)→Gln | Alternative | 8 | 4022.88 | −28.69 | 22.58 |
| TAA(Stp)→Tyr; TGA(Stp)→Trp | Alternative | 2 | 4041.92 | −28.35 | 30.50 |
| TAA(Stp)→Tyr; TAG(Stp)→Tyr | Alternative | 1 | 4116.26 | −27.04 | 59.50 |
| TAA(Stp)→Glu; TGA(Stp)→Trp | Alternative | 1 | 4381.92 | −22.33 | 83.33 |
| TAG(Stp)→Glu; TGA(Stp)→Trp | Alternative | 1 | 4381.92 | −22.33 | 83.33 |
| TAA(Stp)→Glu; TAG(Stp)→Glu | Alternative | 2 | 4692.9 | −16.81 | 98.08 |
| TAG(Stp)→Asp; TGA(Stp)→Asp | Max | 0 | 5032.04 | −10.80 | 100 |

The double codon reassignments were compared with the selected reassignments responsible for the biggest drop (min) and the biggest increase (max) in the values of the function *F*. *B* is the percent of theoretical codes recoding two stop codons to sense codons and having the *F* values smaller than the code with the given reassignment. Other columns as in Table 3.

**Table 7**
The triple codon reassignments observed in alternative genetic codes.

| Reassignment | Type | *R* | *F* | *Pd* | *B* |
|---|---|---|---|---|---|
| TAA(Stp)→Gly; TAG(Stp)→Ser; TGA(Stp)→Ala | min nonsense | 0 | 2861.46 | −49.28 | 0 |
| TAA(Stp)→Gln; TAG(Stp)→Gln; TGA(Stp)→Trp | alternative | 2 | 2945.12 | −47.8 | 0 |
| TAA(Stp)→Glu; TAG(Stp)→Glu; TGA(Stp)→Trp | alternative | 1 | 3697.04 | −34.47 | 5.40E-05 |
| AGA(Arg)→Ser; TAA(Stp)→Tyr; TGA(Stp)→Trp | alternative | 2 | 3999.04 | −29.11 | 0.04 |
| AGG(Arg)→Ser; TAA(Stp)→Tyr; TGA(Stp)→Trp | alternative | 2 | 4001.6 | −29.07 | 0.04 |
| TGA(Stp)→Trp; TAA(Stp)→Tyr; ATA(Ile)→Met | alternative | 1 | 4025.28 | −28.65 | 0.06 |
| AAA(Lys)→Asn; TAA(Stp)→Tyr; TGA(Stp)→Trp | alternative | 1 | 4036.74 | −28.45 | 0.07 |
| AGA(Arg)→Ser; AGG(Arg)→Ser; TGA(Stp)→Trp | alternative | 1 | 4710.92 | −16.49 | 0.62 |
| AGA(Arg)→Gly; AGG(Arg)→Gly; TGA(Stp)→Trp | alternative | 1 | 4718.92 | −16.35 | 0.64 |
| AGA(Arg)→Ser; ATA(Ile)→Met; TGA(Stp)→Trp | alternative | 2 | 4747.4 | −15.85 | 0.74 |
| AGG(Arg)→Ser; ATA(Ile)→Met; TGA(Stp)→Trp | alternative | 2 | 4747.4 | −15.85 | 0.74 |
| AGA(Arg)→Gly; ATA(Ile)→Met; TGA(Stp)→Trp | alternative | 1 | 4748.84 | −15.82 | 0.75 |
| AGG(Arg)→Gly; ATA(Ile)→Met; TGA(Stp)→Trp | alternative | 1 | 4748.84 | −15.82 | 0.75 |
| AAA(Lys)→Asn; AGA(Arg)→Ser; TGA(Stp)→Trp | alternative | 2 | 4757.4 | −15.67 | 0.79 |
| AAA(Lys)→Asn; AGG(Arg)→Ser; TGA(Stp)→Trp | alternative | 2 | 4760.6 | −15.61 | 0.81 |
| CTG(Leu)→Thr; TGA(Stp)→Trp; ATA(Ile)→Met | alternative | 1 | 4766.98 | −15.50 | 0.86 |
| CTA(Leu)→Thr; TGA(Stp)→Trp; ATA(Ile)→Met | alternative | 1 | 4766.98 | −15.50 | 0.86 |
| CTC(Leu)→Thr; TGA(Stp)→Trp; ATA(Ile)→Met | alternative | 1 | 4770.38 | −15.44 | 0.88 |
| CTT(Leu)→Thr; TGA(Stp)→Trp; ATA(Ile)→Met | alternative | 1 | 4770.38 | −15.44 | 0.88 |
| AAA(Lys)→Asn; ATA(Ile)→Met; TGA(Stp)→Trp | alternative | 1 | 4784.44 | −15.19 | 1.00 |
| AGA(Arg)→Ser; AGG(Arg)→Lys; TGA(Stp)→Trp | alternative | 1 | 4839.88 | −14.21 | 1.75 |
| AGA(Arg)→Ser; AGG(Arg)→Ser; TAA(Stp)→Tyr | alternative | 2 | 5066.9 | −10.18 | 5.81 |
| AGG(Arg)→Ser; TAA(Stp)→Tyr; ATA(Ile)→Met | alternative | 1 | 5084.18 | −9.88 | 6.07 |
| AAA(Lys)→Asn; AGG(Arg)→Ser; TAA(Stp)→Tyr | alternative | 1 | 5095.64 | −9.68 | 6.24 |
| AGA(Arg)→Ser; TAA(Stp)→Tyr; ATA(Ile)→Met | alternative | 1 | 5103.38 | −9.54 | 6.36 |
| AAA(Lys)→Asn; AGA(Arg)→Ser; TAA(Stp)→Tyr | alternative | 1 | 5111.64 | −9.39 | 6.48 |
| GAR(Glu)→Asp; GAT(Asp)→Gln; CAC(Asp)→Gln | min missense | 0 | 5116.3 | −9.31 | 6.55 |
| AGA(Arg)→Ser; AGG(Arg)→Ser; ATA(Ile)→Met | alternative | 2 | 5553.14 | −1.57 | 14 |
| AGA(Arg)→Gly; AGG(Arg)→Gly; ATA(Ile)→Met | alternative | 1 | 5556.66 | −1.5 | 14.1 |
| AAA(Lys)→Asn; AGA(Arg)→Ser; AGG(Arg)→Ser | alternative | 2 | 5563.14 | −1.39 | 14.31 |
| AAA(Lys)→Asn; AGG(Arg)→Ser; ATA(Ile)→Met | alternative | 1 | 5581.22 | −1.07 | 14.97 |
| AAA(Lys)→Asn; AGA(Arg)→Ser; ATA(Ile)→Met | alternative | 1 | 5599.78 | −0.74 | 15.91 |
| AGA(Arg)→Stp; ATA(Ile)→Met; TGA(Stp)→Trp | alternative | 1 | 5883.32 | 4.29 | 54.78 |
| AGG(Arg)→Stp; ATA(Ile)→Met; TGA(Stp)→Trp | alternative | 1 | 5883.32 | 4.29 | 54.78 |
| AGA(Arg)→Stp; AGG(Arg)→Stp; TGA(Stp)→Trp | alternative | 1 | 6726.72 | 19.24 | 89.9 |
| AGA(Arg)→Stp; AGG(Arg)→Stp; ATA(Ile)→Met | alternative | 1 | 7320.02 | 29.75 | 98.53 |
| ATY(Ile)→Asp; TTY(Phe)→Asp; CTA(Leu)→Asp | max missense | 0 | 8353.46 | 48.07 | 99.98 |
| TTY(Phe)→Stp; TCY(Ser)→Stp; CCY(Pro)→Stp | max nonsense | 0 | 9112.2 | 61.52 | 100 |

The triple codon reassignments were compared with the selected reassignments responsible for the biggest drop (min) and the biggest increase (max) in the values of the function *F*. Other columns as in Table 3.

**Table 8**
The triple codon reassignments recoding a stop codon to an amino acid in alternative genetic codes.

| Reassignment | Type | *R* | *F* | *Pd* | *B* |
|---|---|---|---|---|---|
| TAA(Stp)→Gly; TAG(Stp)→Ser; TGA(Stp)→Ala | min | 0 | 2861.46 | −49.28 | 0 |
| TAA(Stp)→Gln; TAG(Stp)→Gln; TGA(Stp)→Trp | alternative | 2 | 2945.12 | −47.80 | 9.83 |
| TAA(Stp)→Glu; TAG(Stp)→Glu; TGA(Stp)→Trp | alternative | 1 | 3697.04 | −34.47 | 95.38 |
| TAA(Stp)→Leu/Ile; TAG(Stp)→Asp; TGA(Stp)→Asp | max | 0 | 4326.56 | −23.31 | 100 |

The triple codon reassignments were compared with the selected reassignments responsible for the biggest drop (min) and the biggest increase (max) in the values of the function *F*. *B* is the percentage of the theoretical codes recoding three stop codons to amino acids and having the *F* values smaller than the code with the given reassignment. Other columns as in Table 3.

applied the PAM74-100 matrix to measure the differences between amino acids, however, such an approach was criticized by Di Giulio (2001) and Gilis et al. (2001) because this mutation-selection matrix already includes the structure of the genetic code. Sammet et al. (2010) analysed seven alternative codes in terms of protein misfolding in the case of three proteins with known three-dimensional structure and found one code better than the SGC. The authors proposed that the alternative genetic codes evolved mainly neutrally or nearly-neutrally but they did not rule out any adaptation in some cases. Kurnaz et al. (2010) found two out of 15 alternative codes better than the SGC using various reduced amino acid alphabets based on Miyazawa-Jernigan potentials and the BLOSUM matrix. The optimality of the codes was tested on genes which products are important in some biological processes. However, the BLOSUM matrix creates the same problem as the PAM matrix because it was also derived from protein sequences subjected to mutation and selection associated with the genetic code structure. Moreover, the genes used in the study may not represent the ancestral genes in organisms in which a given alternative genetic code evolved.

Additionally, all the above-mentioned analyses did not consider nonsense mutations, which are important for protein-coding sequences. This type of mutations was included only by Morgens and Cavalcanti (2013), who applied a cost matrix based on the change in the folding free energy in a set of protein structures (Gilis et al., 2001). Using these assumptions, they found that 13 out of 16 alternative codes are more robust in terms of error minimization than the SGC. Here, we significantly extended their and our former approaches (Błażej et al., 2018). We compared more natural codes with all the possible theoretical genetic codes that differed from the SGC in one, two or three codon changes. We also assessed the optimality of individual codon reassignments.

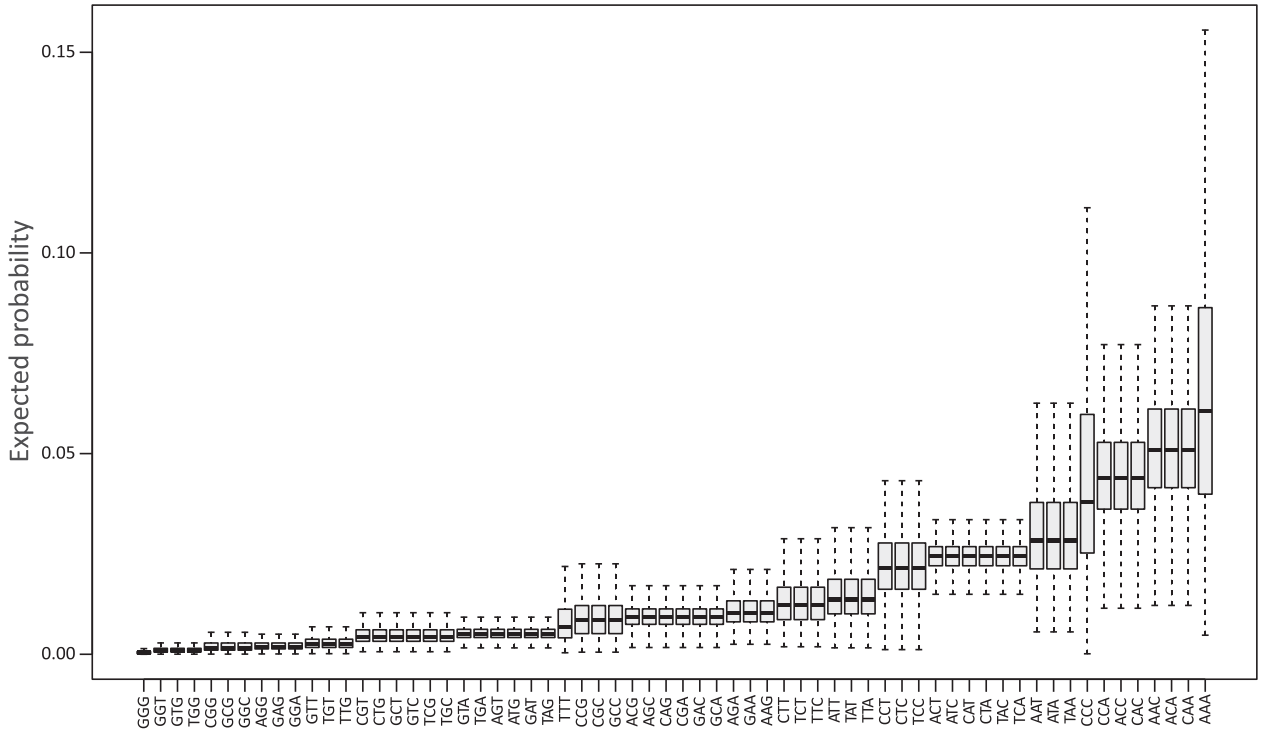


**Fig. 3.** Expected probability of codon occurrence in vertebrate mitochondrial genomes calculated based on the nucleotide composition in the four-fold degenerated sites. The thick line indicates the median, the box shows the quartile range and the whiskers denote the range without outliers.

Our new study demonstrated that 18 out of 21 alternative codes minimize the consequences of point mutations or mistranslations more effectively than the standard genetic code, even up to 48%. The alternative codes also did well in the comparison with all the possible theoretical codes differing from the SGC by up to three codon reassignments. For example, 13 alternative codes maintaining at least one stop codon were outperformed by only up to 4% of the theoretical codes.

The reassignments of stop codons to sense codons and the decrease or even elimination of the stop codons from the genetic codes are responsible for the biggest improvement of the code robustness. It might seem obvious because for the substitutions involving these codons have assigned the largest possible difference observed between the polarity values of amino acids, just like for the most disadvantageous missense mutation. Such nonsense mutations are known to be especially harmful because they result in a premature stop codon and usually non-functional protein products. Therefore, the assumption made in our approach and also by Morgens and Cavalcanti (2013) on the magnitude of the error caused by the nonsense mutations is reasonable. The most beneficial nonsense changes reassign the stop codons to sense codons encoding amino acids characterized by a moderate polarity, i.e. Ala = 7 and Gly = 7.9, as well as Ser = 7.5 in the case of the triple theoretical reassignments. The reassignment to Ala is present only in the theoretical codes but the reassignment to Gly can be found also in the alternative codes.

Our analyses showed that the good performance of the alternative codes results not only from the decrease in number or complete disappearance of the stop codons but also from the missense reassignments. The most beneficial reassignments of such type present in the alternative codes involve changes in the meaning of codons from amino acids with the extreme polarity values (Arg = 9.1, Leu = 4.9) to those with smaller or moderate polarity values, i.e. Ser = 7.5, Gly = 7.9, Thr = 6.6. In the case of the theoretical codes, the most beneficial reassignments are the ones of the codons GAN, previously coding for amino acids with extreme polarity, i.e. Asp = 13 and Glu = 12.5, now ascribed to amino acids with smaller polarity values: Arg = 9.1 and Gln = 8.6. Then, many potential substitutions in the first codon position of the GAN codons are not so deleterious because the mutated codons AAN, CAN and TAN encode amino acids with polarity values similar (8.4–10.1) to those of the newly ascribed amino acids, i.e. Arg and Gln. On the other hand, the most unfavourable missense reassignment ascribes Asp with the highest polarity value 13 to the codons that can be easily mutated into codons encoding amino acids with small polarity value, i.e. Phe = 5. However, such assignments were not found in the alternative codes.

Three alternative codes turned out less robust than the SGC. One of them is the Alternative Yeast Nuclear Code, which reassigns CTG codon from Leu to Ser, and as a result decreases the robustness of the code in terms of polarity. However, this disadvantageous effect in this case is likely overridden by the benefits associated with the accumulation of misfolded and/or aggregated proteins leading to greater phenotypic diversity and better adaptation to more demanding and changing environment (Gomes et al., 2007; Santos et al., 1999).

Another less optimal code is the *Thraustochytrium* Mitochondrial Code in which TTA is an additional stop codon instead of encoding Leu. Since the annotated genomic sequences and other data about this organism are not available, it is not possible to assess the importance of this reassignment. A great amount of data is, however, available for the vertebrate mitochondrial genomes whose code is the least optimal one compared with the SGC. This alternative code has two beneficial changes TGA(Stp)→Trp and ATA(Ile)→Met but also two unfavourable ones, which generate additional stop codons from two arginine codons AGA and AGG (Osawa et al., 1989). The two latter reassignments can be in fact deleterious when such codons appear in protein coding sequences because they result in truncation of the protein products. We assessed the probability of occurrence of such codons in the mitochondrial genomes of vertebrates due to the mutational pressure, which was approximated by the nucleotide composition of

four-fold degenerated sites of protein coding sequences. This probability seems relatively low mainly because of the low rate of substitutions generating guanine. The median nucleotide composition of these sites in the vertebrate mitochondrial genomes is: A = 0.393, C = 0.336, G = 0.070 and T = 0.189. In consequence, the median of the expected probability of AGG and AGA occurrences calculated for all the vertebrate mitochondrial genomes is 0.0019 and 0.0102, respectively. It is much below 0.0156, i.e. the probability of uniform codon generation. Thereby, AGG and AGA are one of the least frequently generated codons in the genomes (Fig. 3). It cannot be ruled out that these probabilities in the ancestor of vertebrates was even lower because the two earliest diverged lineages of vertebrates, i.e. Cyclostomata and Chondrichthyes, are characterized by lower probabilities of occurrence of these codons: AGG = 0.0013 and AGA = 0.0090, and AGG = 0.0008 and AGA = 0.0071, respectively. In a sister lineage to vertebrates, cephalochordates, AGG is even absent. The low probability of occurrence of these codons could explain why such changes in this genetic code were accepted and were not so deleterious.

## 5. Conclusions

Although our results indicate that the alternative codes are quite robust against amino acid replacements compared with the theoretical ones, it does not automatically imply that their reassignments were directly selected to minimize the mutational errors. The codon changes could have been a by-product of processes that recoded rare stop codons to sense codons, e.g. by the codon disappearance mechanism associated with AT-biased mutational pressure (Osawa and Jukes, 1989; Osawa and Jukes, 1995; Sengupta et al., 2007). Missense reassignments can be explained by taking over tRNAs charged by a similar amino acid and posttranscriptional base modifications or point mutations in the tRNAs. Such reassignments could have happened mainly by the mechanisms of ambiguous intermediate (Schultz and Yarus, 1994; Schultz and Yarus, 1996; Sengupta et al., 2007) or unassigned codon (Sengupta and Higgs, 2005; Sengupta et al., 2007). Nevertheless, the significant excess of the beneficial codon changes in the alternative codes implies that at least some of them could have been positively selected or the benefits could have helped with their fixation even if they arose neutrally. The advantageous reassignments are especially important for the organelle genomes which are characterized by high mutational pressures (Lynch et al., 2006). Under the conditions of a strong mutational pressure, the property of minimizing the consequences of mutations by the genetic code have a better chance to be positively selected.

In the light of the presented analyses, the standard genetic code turns out much less robust against point mutations associated with amino acid replacements than most alternative genetic codes. It is also surpassed in the robustness by many theoretical codes differing from it by up to three reassignments although the SGC is slightly closer to the codes minimizing the consequences of amino acid replacements rather than to those maximizing them. For example, even if we assume the arrangement of stop codons as in the SGC, there are almost 23 million codes better than the SGC and this code can be improved by more than 9% by only three missense reassignments. Many other codon changes that can also decrease the error minimization of the SGC are in fact realized in the natural alternative genetic codes.

These findings suggest that the minimization of the effects of point mutations or mistranslation was not the main force behind the evolution of the SGC as it is claimed by the adaptive hypothesis. Although the SGC outperformes most randomly-generated codes (Freeland and Hurst, 1998; Freeland et al., 2000; Gilis et al., 2001; Goodarzi et al., 2004; Haig and Hurst, 1991), other analyses, using optimization algorithms, have shown that the SGC is not as optimized as previously thought (Błażej et al., 2018; Błażej et al., 2016; Judson and Haydon, 1999; Massey, 2008; Novozhilov et al., 2007; Santos and Monteagudo, 2011; Wnętrzak et al., 2018). It was proposed that the increasing diversity of amino acids subsequently added to the expanding code was more important (Higgs, 2009; Sengupta and Higgs, 2015; Weberndorfer et al., 2003). This addition associated with the duplication of genes for tRNAs and aminoacyl-tRNA synthetases charging similar amino acids could have led to some optimization of the SGC, which is not, however, as good as it could be expected based on the adaptive hypothesis (Cavalcanti et al., 2000, 2004; Koonin, 2017; Koonin and Novozhilov, 2017; Massey, 2016; Stoltzfus and Yampolsky, 2007). The process of code expansion was also exerted by the evolution of biosynthetic pathways of amino acids (Di Giulio, 1997, 1999, 2004, 2008, 2016, 2017, 2018; Facchiano and Di Giulio, 2018; Wong, 1975; Wong et al., 2016), while the reduction of mutation errors could have been adjusted by the direct optimization of the mutational pressure around the fixed genetic code (Błażej et al., 2015, 2017; Dudkiewicz et al., 2005; Mackiewicz et al., 2008). On the other hand, after the establishment of the standard code, its variants were subjected to other mechanisms of evolution under a stabilizing selection (Sengupta and Higgs, 2015). Therefore, neutral or beneficial codon reassignments were only accepted if they did not disrupt the already well-adapted system. The study of the properties of the genetic code variants can be helpful in the designing of alternative genetic codes for artificially modified organisms in the framework of synthetic biology and xenobiology (Chin, 2014; Schmidt, 2010; Schmidt and de Lorenzo, 2012; Xie and Schultz, 2006).

## Acknowledgements

We are very grateful for the insightful comments and remarks of two anonymous reviewers, which significantly improved the manuscript.

## Formatting of funding sources

This work was supported by the National Science Centre, Poland (Narodowe Centrum Nauki, Polska) under Grant number 2017/27/N/NZ2/00403.

## References

Abascal, F., Posada, D., Zardoya, R., 2012. The evolution of the mitochondrial genetic code in arthropods revisited. Mitochondrial DNA 23, 84–91.
Andersson, S.G.E., Kurland, C.G., 1991. An extreme codon preference strategy - codon reassignment. Mol. Biol. Evol. 8, 530–544.
Andersson, S.G.E., Kurland, C.G., 1995. Genomic evolution drives the evolution of the translation system. Biochem. Cell Biol. 73, 775–787.
Bender, A., Hajieva, P., 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. U. S. A. 105, 16496–16501.
Błażej, P., Wnętrzak, M., Mackiewicz, D., Mackiewicz, P., 2018. Optimization of the standard genetic code according to three codon positions using an evolutionary algorithm. PLoS One 13, e0201715.
Błażej, P., Wnętrzak, M., Mackiewicz, P., 2016. The role of crossover operator in evolutionary-based approach to the problem of genetic code optimization. BioSystems 150, 61–72.
Błażej, P., Wnętrzak, M., Mackiewicz, P., 2018. The importance of changes observed in the alternative genetic codes. In: Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 4: Bioinformatics, pp. 154–159.
Błażej, P., Miasojedow, B., Grabińska, M., Mackiewicz, P., 2015. Optimization of mutation pressure in relation to properties of protein-coding sequences in bacterial genomes. PLoS One 10, e0130411.
Błażej, P., Mackiewicz, D., Grabińska, M., Wnętrzak, M., Mackiewicz, P., 2017. Optimization of amino acid replacement costs by mutational pressure in bacterial genomes. Sci. Rep. 7, 1061.
Boore, J.L., Brown, W.M., 1994. Complete DNA sequence of the mitochondrial genome of the black chiton, Katharina tunicata. Genetics 138, 423–443.

Bove, J.M., 1993. Molecular features of mollicutes. Clin. Infect. Dis. 17 (Suppl 1), S10–S31.
Bracht, J.R., Fang, W., Goldman, A.D., Dolzhenko, E., Stein, E.M., Landweber, L.F., 2013. Genomes on the edge: programmed genome instability in ciliates. Cell 152, 406–416.
Campbell, J.H., O'Donoghue, P., Campbell, A.G., Schwientek, P., Sczyrba, A., Woyke, T., Soll, D., Podar, M., 2013. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc. Natl. Acad. Sci. U. S. A. 110, 5540–5545.
Cavalcanti, A.R., Neto, B.D., Ferreira, R., 2000. On the classes of aminoacyl-tRNA synthetases and the error minimization in the genetic code. J.Theor, Biol, 204, 15–20.
Cavalcanti, A.R.O., Leite, E.S., Neto, B.B., Ferreira, R., 2004. On the classes of aminoacyl-tRNA synthetases, amino acids and the genetic code. Origin, Life Evol. Biosphere 34, 407–420.
Chin, J.W., 2014. Expanding and reprogramming the genetic code of cells and animals. Annu. Rev. Biochem. 83, 379–408.
Clark-Walker, G.D., Weiller, G.F., 1994. The structure of the small mitochondrial DNA of Kluyveromyces thermotolerans is likely to reflect the ancestral gene order in fungi. J. Mol. Evol. 38, 593–601.
Crozier, R.H., Crozier, Y.C., 1993. The mitochondrial genome of the honeybee Apis mellifera: complete sequence and genome organization. Genetics 133, 97–117.
Del Cortona, A., Leliaert, F., Bogaert, K.A., Turmel, M., Boedeker, C., Janouskovec, J., Lopez-Bautista, J.M., Verbruggen, H., Vandepoele, K., De Clerck, O., 2017. The plastid genome in cladophorales green algae is encoded by hairpin chromosomes. Curr. Biol. 27, 3771–3782 e6.
Di Giulio, M., 1997. On the origin of the genetic code. J. Theor. Biol. 187, 573–581.
Di Giulio, M., 1999. The coevolution theory of the origin of the genetic code. J. Mol. Evol. 48, 253–255.
Di Giulio, M., 2001. The origin of the genetic code cannot be studied using measurements based on the PAM matrix because this matrix reflects the code itself, making any such analyses tautologous. J. Theor. Biol. 208, 141–144.
Di Giulio, M., 2004. The coevolution theory of the origin of the genetic code. Phys. Life Rev. 1, 128–137.
Di Giulio, M., 2008. An extension of the coevolution theory of the origin of the genetic code. Biol. Direct 3, 37.
Di Giulio, M., 2016. The lack of foundation in the mechanism on which are based the physico-chemical theories for the origin of the genetic code is counterposed to the credible and natural mechanism suggested by the coevolution theory. J. Theor. Biol. 399, 134–140.
Di Giulio, M., 2017. Some pungent arguments against the physico-chemical theories of the origin of the genetic code and corroborating the coevolution theory. J. Theor. Biol. 414, 1–4.
Di Giulio, M., 2018. A discriminative test among the different theories proposed to explain the origin of the genetic code: the coevolution theory finds additional support. BioSystems 169-170, 1–4.
Dudkiewicz, A., Mackiewicz, P., Nowicka, A., Kowalezuk, M., Mackiewicz, D., Polak, N., Smolarczyk, K., Banaszak, J., Dudek, M.R., Cebrat, S., 2005. Correspondence between mutation and selection pressure and the genetic code degeneracy in the gene evolution. Fut. Gener. Comput. Syst. 21, 1033–1039.
Facchiano, A., Di Giulio, M., 2018. The genetic code is not an optimal code in a model taking into account both the biosynthetic relationships between amino acids and their physicochemical properties. J. Theor. Biol..
Freeland, S.J., Hurst, L.D., 1998. The genetic code is one in a million. J. Mol. Evol. 47, 238–248.
Freeland, S.J., Knight, R.D., Landweber, L.F., Hurst, L.D., 2000. Early fixation of an optimal genetic code. Mol. Biol. Evol. 17, 511–518.
Gilis, D., Massar, S., Cerf, N.J., Rooman, M., 2001. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. Genome Biol. 2 research 0049.1–0049.12.
Gomes, A.C., Miranda, I., Silva, R.M., Moura, G.R., Thomas, B., Akoulitchev, A., Santos, M.A., 2007. A genetic code alteration generates a proteome of high diversity in the human pathogen Candida albicans. Genome Biol. 8.
Goodarzi, H., Nejad, H.A., Torabi, N., 2004. On the optimality of the genetic code, with the consideration of termination codons. BioSystems 77, 163–173.
Haig, D., Hurst, L.D., 1991. A quantitative measure of error minimization in the genetic code. J. Mol. Evol. 33, 412–417.
Heaphy, S.M., Mariotti, M., Gladyshev, V.N., Atkins, J.F., 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.
Higgs, P.G., 2009. A four-column theory for the origin of the genetic code: tracing the evolutionary pathways that gave rise to an optimized code. Biol. Direct 4, 16.
Hoffman, D.C., Anderson, R.C., Dubois, M.L., Prescott, D.M., 1995. Macronuclear gene-sized molecules of hypotrichs. Nucleic Acids Res. 23, 1279–1283.
Janouskovec, J., Sobotka, R., Lai, D.H., Flegontov, P., Konik, P., Komenda, J., Ali, S., Prasil, O., Pain, A., Obornik, M., Lukes, J., Keeling, P.J., 2013. Split photosystem protein, linear-mapping topology, and growth of structural complexity in the plastid genome of Chromera velia. Mol. Biol. Evol. 30, 2447–2462.
Judson, O.P., Haydon, D., 1999. The genetic code: what is it good for? An analysis of the effects of selection pressures on genetic codes. J. Mol. Evol. 49, 539–550.
Knight, R.D., Landweber, L.F., Yarus, M., 2001a. How mitochondria redefine the code. J. Mol. Evol. 53, 299–313.
Knight, R.D., Freeland, S.J., Landweber, L.F., 2001b. Rewiring the keyboard evolvability of the genetic code. Nat. Rev. Genet. 2, 49–58.
Koonin, E.V., 2017. Frozen accident pushing 50: stereochemistry, expansion, and chance in the evolution of the genetic code. Life (Basel) 7, 22.
Koonin, E.V., Novozhilov, A.S., 2017. Origin and evolution of the universal genetic code. Annu. Rev. Genet. 51, 45–62.
Kurnaz, M.L., Bilgin, T., Kurnaz, I.A., 2010. Certain non-standard coding tables appear to be more robust to error than the standard genetic code. J. Mol. Evol. 70, 13–28.
Lang-Unnasch, N., Aiello, D.P., 1999. Sequence evidence for an altered genetic code in the Neospora caninum plastid. Int. J. Parasitol. 29, 1557–1562.
Lim, P.O., Sears, B.B., 1992. Evolutionary relationships of a plant-pathogenic *Mycoplasma*-like organism and *Acholeplasma laidlawii* deduced from 2 ribosomal protein gene sequences. J. Bacteriol. 174, 2606–2611.
Lynch, M., Koskella, B., Schaack, S., 2006. Mutation pressure and the evolution of organelle genomic architecture. Science 311, 1727–1730.
Mackiewicz, P., Biecek, P., Mackiewicz, D., Kiraga, J., Baczkowski, K., Sobczynski, M., Cebrat, S., 2008. Optimisation of asymmetric mutational pressure and selection pressure around the universal genetic code. In: Computational Science - ICCS 2008, 5103, pp. 100–109.
Massey, S.E., 2008. A neutral origin for error minimization in the genetic code. J. Mole. Evol. 67, 510–516.
Massey, S.E., 2016. The neutral emergence of error minimized genetic codes superior to the standard genetic code. J. Theor. Biol. 408, 237–242.
Massey, S.E., Garey, J.R., 2007. A comparative genomics analysis of codon reassignments reveals a link with mitochondrial proteome size and a mechanism of genetic code change via suppressor tRNAs. J. Mole. Evol. 64, 399–410.
McCutcheon, J.P., McDonald, B.R., 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.
Mochizuki, K., 2010. DNA rearrangements directed by non-coding RNAs in ciliates. Wiley Interdiscip. Rev. 1, 376–387.
Morgens, D.W., Cavalcanti, A.R., 2013. An alternative look at code evolution: using non-canonical codes to evaluate adaptive and historic models for the origin of the genetic code. J. Mol. Evol. 76, 71–80.
Muhlhausen, S., Findeisen, P., Plessmann, U., Urlaub, H., Kollmar, M., 2016. A novel nuclear genetic code alteration in yeasts and the evolution of codon reassignment in eukaryotes. Genome Res. 26, 945–955.
Novozhilov, A.S., Wolf, Y.I., Koonin, E.V., 2007. Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. Biol. Direct 2, 24.
Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T.G., Landweber, L.F., 2008. RNA-mediated epigenetic programming of a genome-rearrangement pathway. Nature 451, 153.
Osawa, S., Jukes, T.H., 1989. Codon reassignment (codon capture) in evolution. J. Mol. Evol. 28, 271–278.
Osawa, S., Jukes, T.H., 1995. On codon reassignment. J. Mol. Evol. 41, 247–249.
Osawa, S., Ohama, T., Jukes, T.H., Watanabe, K., 1989. Evolution of the mitochondrial genetic code. I. Origin of AGR serine and stop codons in metazoan mitochondria. J. Mol. Evol. 29, 202–207.
Panek, T., Zihala, D., Sokol, M., Derelle, R., Klimes, V., Hradilova, M., Zadrobilkova, E., Susko, E., Roger, A.J., Cepicka, I., Elias, M., 2017. Nuclear genetic codes with a different meaning of the UAG and the UAA codon. BMC Biol. 15, 8.
R_Core_Team, 2017. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/.
Sammet, S.G., Bastolla, U., Porto, M., 2010. Comparison of translation loads for standard and alternative genetic codes. BMC Evol. Biol. 10, 178.
Sanchez-Silva, R., Villalobo, E., Morin, L., Torres, A., 2003. A new noncanonical nuclear genetic code: translation of UAA into glutamate. Curr. Biol. 13, 442–447.
Santos, J., Monteagudo, A., 2010. Study of the genetic code adaptability by means of a genetic algorithm. J. Theor. Biol. 264, 854–865.
Santos, J., Monteagudo, A., 2011. Simulated evolution applied to study the genetic code optimality using a model of codon reassignments. BMC Bioinform. 12, 56.
Santos, M.A.S., Keith, G., Tuite, M.F., 1993. Nonstandard translational events in Candida-Albicans mediated by an unusual seryl-transfer RNA with a 5′-Cag-3′ (Leucine) anticodon. Embo J. 12, 607–616.
Santos, M.A.S., Cheesman, C., Costa, V., Moradas-Ferreira, P., Tuite, M.F., 1999. Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in Candida spp. Mol. Microbiol. 31, 937–947.
Schmidt, M., 2010. Xenobiology: A new form of life as the ultimate biosafety tool. Bioessays 32, 322–331.
Schmidt, M., de Lorenzo, V., 2012. Synthetic constructs in/for the environment: managing the interplay between natural and engineered Biology. Febs Lett. 586, 2199–2206.
Schneider, S.U., Leible, M.B., Yang, X.P., 1989. Strong Homology between the small subunit of ribulose-1,5-bisphosphate carboxylase oxygenase of 2 species of acetabularia and the occurrence of unusual codon usage. Mol. Gen. Genet. 218, 445–452.
Schultz, D.W., Yarus, M., 1994. Transfer-RNA mutation and the malleability of the genetic code. J. Mol. Biol. 235, 1377–1380.
Schultz, D.W., Yarus, M., 1996. On malleability in the genetic code. J. Mol. Evol. 42, 597–601.
Sengupta, S., Higgs, P.G., 2005. A unified model of codon reassignment in alternative genetic codes. Genetics 170, 831–840.
Sengupta, S., Higgs, P.G., 2015. Pathways of genetic code evolution in ancient and modern organisms. J. Mol. Evol. 80, 229–243.
Sengupta, S., Yang, X., Higgs, P.G., 2007. The mechanisms of codon reassignments in mitochondrial genetic codes. J. Mol. Evol. 64, 662–688.

Sonneborn, T.M., 1965. Degeneracy of the genetic code: extent, nature, and genetic implications. In: Bryson, V., Vogel, H.J. (Eds.), Evolving Genes and Proteins. Academic Press, New York, pp. 377–397.
Stoltzfus, A., Yampolsky, L.Y., 2007. Amino acid exchangeability and the adaptive code hypothesis. J. Mol. Evol. 65, 456–462.
Swire, J., Judson, O.P., Burt, A., 2005. Mitochondrial genetic codes evolve to match amino acid requirements of proteins. J. Mol. Evol. 60, 128–139.
Weberndorfer, G., Hofacker, I.L., Stadler, P.F., 2003. On the evolution of primitive genetic codes. Origin. Life Evol. Biospheres 33, 491–514.
Wnętrzak, M., Błażej, P., Mackiewicz, D., Mackiewicz, P., 2018. The optimality of the standard genetic code assessed by an eight-objective evolutionary algorithm. BMC Evol. Biol. 18, 192.
Woese, C.R., 1965. On the evolution of the genetic code. Proc. Natl. Acad. Sci. U. S. A. 54, 1546–1552.
Woese, C.R., 1973. Evolution of the genetic code. Naturwissenschaften 60, 447–459.
Wong, J.T., 1975. A co-evolution theory of the genetic code. Proc. Natl. Acad. Sci. U. S. A. 72, 1909–1912.
Wong, J.T., Ng, S.K., Mat, W.K., Hu, T., Xue, H., 2016. Coevolution theory of the genetic code at age forty: pathway to translation and synthetic life. Life (Basel) 6.
Xie, J.M., Schultz, P.G., 2006. Innovation: a chemical toolkit for proteins - an expanded genetic code. Nat. Rev. Mol. Cell Biol. 7, 775–782.
Yokobori, S., Watanabe, Y., Oshima, T., 2003. Mitochondrial genome of Ciona savignyi (Urochordata, Ascidiacea, Enterogona): comparison of gene arrangement and tRNA genes with Halocynthia roretzi mitochondrial genome. J. Mol. Evol. 57, 574–587.
Zahonova, K., Kostygov, A.Y., Sevcikova, T., Yurchenko, V., Elias, M., 2016. An unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. Curr. Biol. 26, 2364–2369.
Zufall, R.A., McGrath, C.L., Muse, S.V., Katz, L.A., 2006. Genome architecture drives protein evolution in ciliates. Mol. Biol. Evol. 23, 1681–1687.