Article ID: abb150316104

OPEN ACCESS

Applied Biology & Biotechnology

ISSN: 2320 -7825 (Print); 2320-7833 (Online)

Evolving Genetic Code - Review

B. Singaravelan

UGC-NRCBS
School of Biological Sciences
Madurai Kamaraj University
Madurai 625 021, India. For Correspondence
Email: microgeneticsings@gmail.com
Tel: +91-80126 33278

Received: 24 January 2016 / Accepted:19 March 2016/ Published Online:15 March 2016
http://www.gayathripublishers.com/abb.htm

Citation: Singaravelan, B. 2016. Evolving Genetic Code - Review. *Applied Biology and Biotechnology*, 4(1):14-18.

Abstract
Studies on the periodic behavior of a species are called Genetic code is universal in all living organisms and it is frozen. Any change or alteration in the genetic code might lead to lethality. The canonical genetic code was assigned to 61 sense codons and 3 stop codons. But growing evidence on genetic code reassignments is thawing the Crick's Frozen Accident Theory. These reassignments were observed in both mitochondrial and nuclear genomes and two main theories widely discussed for the codon reassignments are the codon capture theory and the ambiguous intermediate theory. Genetic code evolution may be advantageous or disadvantageous, but the beneficial variation might help the organism to adapt to the environmental needs. This genetic code flexibility opened up a new area of research to introduce non-standard amino acids to study the genotypic variations. In this review, past to the present day possible concepts on genetic code and reassignments have been discussed.

Key Words: Genetic code, reassignments, non-canonical codons, evolution, codon ambiguity.

DNA, the repository of genetic information of an organism is laid down in a linear sequence of four nucleotide bases adenine, thymine, guanine and cytosine in the double stranded DNA which governs all cellular functions. DNA is meticulously transcribed into an intermediate messenger RNA and translated into proteins which execute the tasks necessary for life. The first glimmer of light in the story of the genetic code came when Dounce (1952) proposed the model that the order of nucleotides determines the order of amino acids in polypeptide chains. Initial studies with T4 *rII* mutants, growing support for the existence of an intermediate messenger RNA between DNA and the protein synthesis, *in vitro* translation of polyribonucleotides, adaptor hypothesis and ribosomes paved way for deciphering the genetic code. Non-overlapping triplet nature of the genetic code was cracked by Crick *et al*., (1961) and this assigned the 64 codons to the 20 amino acids.

Discovery of polynucleotide phosphorylase for the preparation of single and mixed polymers (Grunberg-Manago *et al*., 1955) was the initiative for the enzymatic synthesis of oligonucleotides. Nirenberg and Matthaei (1961) using a cell free *E.coli* extracts and polyU, poly phenylalanine was incorporated and confirmed that UUU triplet codes for phenylalanine. This was the first advancement in deciphering the genetic code and it proved that cell free *E.coli* extracts with synthetic template it might be able to synthesize any protein corresponding to the information provided. In a series of articles communicated by Ochoa entitled "synthetic polynucleotides and the amino acid code" synthetic homo and copolymers were synthesized and the amino acid incorporation were studied in *E.coli* system to elucidate the genetic code letters (e.g. Lengyel *et al*., 1961). Advancement to this study lead to the analysis of the complete genetic code was published in a series by Nirenberg and Leder using ribosomal binding technique with the availability of trinucleoside diphosphates, radioactive form of amino acids, good knowledge on tRNA preparation and fractionation, aminoacyl tRNA synthetases, and ribosome (e. g. Nirenberg and Leder, 1964). Specific di- and trinucleotide repeating sequences were synthesized and they were used as a messenger RNA in the cell free amino acid incorporating system and the polypeptides bear one or few amino acids in the repeating pattern (e. g. Nishimura *et al*., 1964). The three nonsense triplet codons, the normal recognition for chain termination in mRNA were deciphered based on suppression analysis (Brenner *et al*., 1965 and 1967). Finally the genetic code was assigned to 64 triplets (codons) in which 61 codons were assigned to 20 canonical amino acids and the rest three acts as termination codon (reviewed in Söll and RajBhandary, 2006).

There are four nucleotide bases for the 20 amino acids and the total number of genetic code should not be less than the amino acid. When two bases are considered for each amino acid (4^2) it forms only 16 doublets which is also less to code for all amino acids. Hence, triplet of bases (4^3) gives 64 triplets three times more than the number needed whereby one amino acid will have multiple codons and this leads to the concept, degeneracy of the genetic code (Hayes, 1998). In 1966, Crick postulated the wobble hypothesis to solve the degeneracy of the genetic code. It states that the first two bases of the codon, pairs with the anticodon canonically but to expand the more recognition of the codons the third base of

the codon and the first base of the anticodon, pairs with relaxed rules and also canonical (Crick, 1966; Murphy and Ramakrishnan, 2004; reviewed by Agris *et al*., 2007). The wobble hypothesis and the discovery of the initiator tRNA further refined the genetic code (reviewed in Söll and RajBhandary, 2006). Initiation of the protein synthesis is marked by $tRNA^{Met}$ and as of date AUG (Met), GUG (Val), UUG (Leu) and AUU (Ile) have been shown to function as initiation codons (McCarthy and Brimacombe, 1994; Sussman *et al*., 1996). During the inception of the genetic code it was suggested that, the genetic code is universal in all living organisms and the universality of the genetic code led Crick to propose the "Frozen Accident Theory", in which he stated that, the genetic code was developed as an evolutionary accidental event and it was frozen. Hence, it cannot evolve further. If there is any alteration, it would change the meaning of the codon and would be lethal (Crick, 1968) which is analogous to reassigning the keyboard, every message typed will have erroneous message (reviewed in Knight *et al*., 2001).

In 1979, the concept of "Frozen Accident Theory" was challenged when reassignment was found in the vertebrate mitochondria where AUA (Ile) codon reassigned to Met and UGA stop codon reassigned to Trp from the standard universal code (reviewed in Osawa *et al*., 1992). Since then, there were lot of reports suggesting that the genetic codon reassignments in nuclear and mitochondrial genomes and they are listed in the NCBI website http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html /index.cgi?chapter=cgencodes. Two main theories widely discussed for the codon reassignments are: the codon capture theory and the ambiguous intermediate theory. The codon capture theory states that a codon and its tRNA carrying the anticodon will disappear from the genome due to GC or AT pressure and if the codon and anticodon re-emerges by genetic drift the tRNA might code for the same amino acid or it may be reassigned to some other amino acid (Osawa and Jukes, 1989). The AT and GC content variations seen among the organisms helped in postulating this theory and it is supported by the unassignment of the CGG codon in *Mycoplasma* spp. (25% GC) and AGA, AUA codons in *Micrococcus* spp. (75% GC). This theory fails to explain the reassignment of AT rich codons in an AT rich genome (Massey, 2015). The ambiguous intermediate theory states that rather than disappearing of the codon during reassignment, a codon is read by cognate and near cognate tRNAs, an intrinsic characteristic feature of translational errors. Over a period of time the near cognate tRNAs displaces the cognate tRNA and gets fixed through selection driven process by codon ambiguity. The recognition of mutant tRNAs by more than one aminoacyl-tRNA synthetase results in mutant tRNAs with double identity; mutations in tRNAs may cause misreading of stop codons and subsequently the release factor may lose their ability in recognizing the stop codon which further facilitates the complete change in the genetic code reassignment (Schultz and Yarus, 1994; 1996) and CUG reassignment from Leu to Ser in fungi strongly favour the hypothesis (Suzuki, *et al*., 1997).

The existence of unassigned, absent and extremely rare codons predicts that the genetic code is evolving and this condition lightens up the fact that these codes are midway in the evolution between disappearance and reassignment. Unassigned and rare codons may be due to the genetic drift between AT or GC pressure and it is also very clear by observing the variation in the GC content between various organisms. In yeast *Torulopsis glabrata*, AT rich mitochondrial genome CGN (N refers to any nucleotide) and the corresponding tRNA coding for arginine is absent and in *S. cerevisiae* these codons are rare, in AT rich mitochondrial genome of *Prototheca wikceramii,* a chlorophyte alga UAG, UGA and CGG codons are unassigned, in *Mycoplasma capricolum* the CGG codon is absent along with the respective tRNA, and in the *Micrococcus luteus* AGA and AUA codons are unassigned (reviewed in Miranda *et al*., 2006; Osawa *et al*., 1992; Ohama *et al*., 2008).

Reassignments were also observed in the termination codons UAG, UAA and UGA (amber, ochre and opal respectively) of several organisms which required for the termination and release of the polypeptide. The termination codons can be read through by the suppressor tRNAs, mutation in ribosomal proteins and release factors, but how the efficient protein synthesizing machinery makes error in adding up the amino acid to the termination codon? In most ciliates the stop codons UAG and UAA are decoded as glutamine which result in only one stop codon and four glutamine codons (Salim *et al*., 2008), in Euplotids UGA encodes for cysteine (Tourancheau *et al*., 1995) whereas in *Blepharisma* and *Colpoda* UGA encodes for tryptophan. Like ciliates, the oxymonad, *S. strix* also bear four glutamine codons, two canonical codons CAA, CAG and two non-canonical codons UAA and UAG and also in the green alga *Acetabularia acetabulum.* In mitochondria of certain green algae, UAG termination codon is decoded as alanine and leucine (Keeling and Leander, 2003; reviewed in Knight *et al*., 2001). In two diplomonads *Hexamita* strains UAA and UAG termination codons were observed in numerous in-frame positions coding for glutamine and the same was not found in other two diplomonads sequenced (Keeling and Doolittle, 1996). In SR1 bacteria and Gracilibacterial strains ACD78 and ACD80 the UGA termination codon was read as glycine (reviewed in Ling *et al*., 2015).

Article ID: abb150316104

OPEN ACCESS

Applied Biology & Biotechnology

ISSN: 2320 -7825 (Print); 2320-7833 (Online)

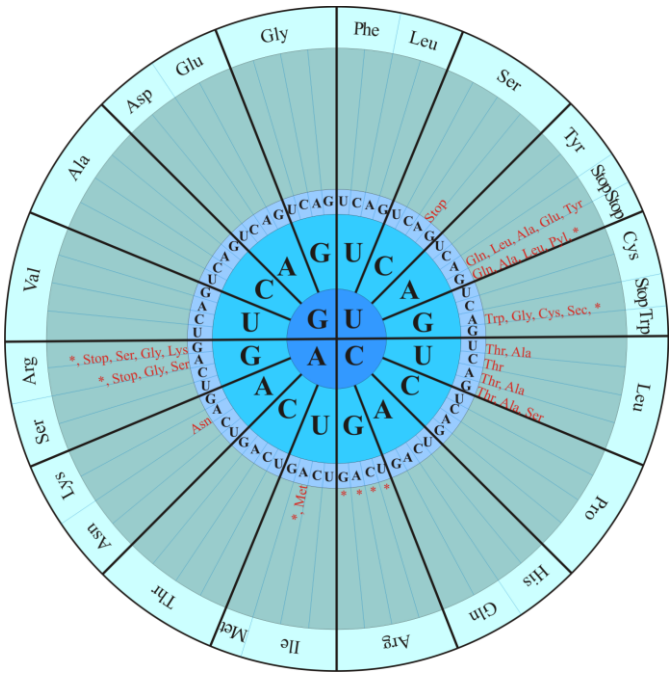

Fig.1: Summary of genetic code reassignments: Standard genetic code is given in black. The reassignments are in red and the unassigned codons are denoted by asterisk symbol.

Strong evidence for confirming the evolvability of the genetic code is the acceptance of 21st and 22nd amino acid, Selenocysteine and Pyrrolysine respectively. To date Sec and Pyl are the two amino acids accepted in the natural genetic code expansion. Selenocysteine is found in all three domains of life, but not in all organisms. UGA tends to play as a termination codon in protein synthesis but when it is present in-frame it is decoded as Sec. Pyrrolysine, the 22nd amino acid was reassigned to the UAG termination codon. It was first discovered in methanogenic archaea and database reports show that it is in 46 microbial species (21 archaea and 25 bacteria) (reviewed in Ling *et al*., 2015; Ambrogelly *et al*., 2007).

Interestingly some of the sense codons were also shown to behave as stop codons. This would happen either due to loss of respective tRNA for rare codons along with modification in the release factor for termination. In mitochondria of *Scenedesmus* UCA (Ser) code acts as a stop codon (reviewed in Knight *et al*., 2001), in vertebrate mitochondrial code AGA and AGG are used as stop codons (Osawa *et al*., 1992). In the yeast mitochondria four leucine codons of the CUN family decoded as threonine instead of leucine by a $tRNA^{Thr}$ with an 8 nucleotide enlarged anticodon loop evolved from yeast mitochondrial $tRNA^{His}$ (Su *et al*., 2011; Miranda *et al*., 2006). In some arthropods arginine codon AGG were found to translate Ser and others use it as Lys or it is an unassigned codon (Abascal *et al*., 2012). In yeast *Candida albicans*, CUG encodes serine rather than leucine (Keeling and Leander, 2003). Suzuki *et al*., (1997) reported that the novel $tRNA^{Ser}$ with the anticodon CAG bear dual specificity, able to charge both leucine (3-5%) and serine (95-97%) *in vivo* in the species *Candida zeylanoides* for the standard genetic code CUG encoding for the leucine. AUA reassigned from Ile to Met independently in animal and yeast mitochondria, in mitochondria of *Ashbya gossypii* CUA and CUU leucine codons are read as alanine (reviewed in Knight *et al*., 2001).

Genome and proteome sequencing and advancement in sequence analyzing tools enhanced the discovery of new genetic code reassignments. In fact the advancement is thawing the frozen genetic code. Is it universal to all three domains of life? How are we going to see the genetic code? Single base change may result in alteration of the phenotype, disappearance of codon and its cognate tRNA as postulated by the codon capture theory might lead to proteome disruption and reduce the fitness of the organism. On the other hand as stated in the ambiguous intermediate theory, codon ambiguity and read through by the near cognate tRNAs might produce new proteins with beneficial functions. Diverse phenotypes in the human pathogen *Candida albicans* were shown due to increase in the genetic code ambiguity (Miranda *et al*., 2007). Without a sudden drastic change in the proteome, slowly new modified proteins evolve and the organism may adapt with new beneficial functions. tRNAs play a major role in decoding the codons by carrying the amino acid to the translational machinery. Apart from the genetic drift, AT or GC pressure, mutations in tRNA, presence of modified nucleotides in the anticodon, minimal number of tRNAs required for decoding with the aid of wobbling hypothesis may also drive a mutational selection pressure for the genetic code reassignments. The existence of unassigned, absent and rare codons strongly favours for the codon capture theory and with the available information one can speculate the possible role of both the theories in the genetic code reassignment. The non-canonical genetic codes have arisen independently at least 10 times in nuclear genomes (reviewed in Söll and RajBhandary, 2006) and at least 24 times in mitochondrial genomes (Swire *et al*., 2005). The existence of unassigned, absent or rare codons implies that there may be some organisms which have lesser than 64 codons. Pezo *et al*., (2013) reassigned the UGG tryptophan codon to histidine and eliminate the tryptophan from *E. coli* and they reduced the amino acids to 19 for the survivability of an organism. This experimental evidence shows that the genetic code can be forced to evolve through natural selection under pressure. All these increasing evidences unveil the flexibility of the genetic code and this flexibility might help us in advancing to a new area of research in engineering the non-standard amino acids into the organism's genetic code to study the beneficial variations. Shedding more light on the genetic code reassignments with genomic and proteomic

approaches might help us to understand novel mechanisms that trigger the genome evolution.

Acknowledgment

I thank all my friends who supported for preparing this manuscript.

References

Abascal, F., Posada, D. and Zardoya, R. 2012. The evolution of the mitochondrial genetic code in arthropods revisited. *Mitochondrial DNA,* 23: 84-91.

Agris, P. F., Vendeix, F. A. P. and Graham, W. D. 2007. tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol*., 366: 1-13.

Ambrogelly, A., Palioura, S. and Söll, D. 2007. Natural expansion of the genetic code. *Nature chemical biology,* 3:29-34.

Brenner, S., Stretton, A. O. W. and Kaplan, S. 1965. Genetic code: The 'nonsense' triplets for chain termination and their suppression. *Nature*, 206: 994-998.

Brenner, S., Barnett, L., Katz, E. R. and Crick, F. H. C. 1967. UGA: A third nonsense triplet in the genetic code. *Nature,* 213: 449-450.

Crick, F. H. C. 1966. Codon-Anticodon pairing: The Wobble Hypothesis. *J. Mol. Biol*., 19: 548-555.

Crick, F. H. C. 1968. The origin of the genetic code. *J. Mol. Biol.,* 38: 367–379.

Crick, F. H., Barnett, L., Brenner, S. and Watts-Tobin, R. J. 1961. General nature of the genetic code for proteins. *Nature,* 192: 1227–1232.

Dounce, A. L. 1952. Duplicating mechanism for peptide chain and nucleic acid synthesis. *Enzymologia* 15: 251–258.

Grunberg-Manago, M., Ortiz, P. J. and Ochoa, S. 1955. Enzymatic synthesis of nucleic acid like polynucleotides. *Science*, 122: 907-910.

Hayes, B. 1998. The invention of the Genetic code. *American scientist,* 86: 8-14.

Keeling, P. J. and Doolittle, W. F. 1996. A non-canonical genetic code in an early diverging eukaryotic lineage. *EMBO J.,* 15:2285-2290.

Keeling, P. J. and Leander, B. S. 2003. Characterisation of a non-canonical genetic code in the oxymonad *Streblomastix strix*. *J. Mol. Bio*., 326: 1337-1349.

Knight, R. D., Freeland, S. J. and Landweber, L. F. 2001. Rewiring the keyboard: Evolvability of the genetic code. *Nat. Rev. Genet.,* 2: 49–58.

Lengyel, P., Speyer, J. F. and Ochoa, S. 1961. Synthetic polynucleotides and the amino acid code. *Proc. Natl. Acad. Sci*. *USA.,* 47: 1936-1942.

Ling, J., O' Donoghue, P. and Söll, D. 2015. Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology. *Nature reviews Microbiology*,13:1-15.

Massey, S. E. 2015. Genetic code evolution reveals the neutral emergence of mutational robustness, and information as an evolutionary constraint. *Life,* 5: 1301-1332.

McCarthy, J. E. and Brimacombe, R. 1994. Prokaryotic translation: the interactive pathway leading to initiation. *Trends Genet*., 10: 402-407.

Miranda, I., Rocha, R., Santos, M. C., Mateus, D. D., Moura, G. R., Carreto, L. and Santos, M. A. 2007. A genetic code alteration is a phenotype diversity generator in the human pathogen *Candida albicans*. *PLoS One*, 2(10):e996.

Miranda, I., Silva, R. and Santos, M. A. S. 2006. Evolution of the genetic code in yeasts. *Yeast*, 23: 203-213.

Murphy, F. V. and Ramakrishanan, V. 2004. Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nature structural and molecular Biology,*11: 1251-1252.

Nirenberg, M. and Leder, P. 1964. RNA codewords and protein synthesis. The effect of trinucleotides upon the binding of sRNA to ribosomes. *Science*, 145: 1399-1407.

Nirenberg, M. W. and Matthaei, J. H. 1961. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci*. *USA.,* 47: 1588-1602.

Nishimura, S., Jacob, T. M. and Khorana, H. G. 1964. Synthetic deoxyribopolynucleotides as templates for ribonucleic acid polymerase: the formation and characterization of a ribopolynucleotide with a repeating trinucleotide sequence. *Proc. Natl. Acad. Sci*. *USA,*. 52: 1494-1501.

Ohama, T., Inagaki, Y., Bessho, Y. and Osawa, S. 2008. Evolving genetic code. *Proc. Jpn. Acad. Ser., B*. 84: 58-74.

Osawa, S. and Jukes, T. H. 1989. Codon reassignment (codon capture) in evolution. *J. Mol. Evol.,* 28: 271–278.

Osawa, S., Jukes, T. H., Watanabe, K. and Muto, A. 1992. Recent evidence for evolution of the genetic code. *Microbiol. Rev.,* 56: 229-264.

Pezo, V., Louis, D., Guèrineau, V., Le Caer, JP., Gaillon, L., Mutzel, R. and Marlière, P. 2013. A metabolic prototype for eliminating tryptophan from the genetic code. *Sci. Rep*. 3:1359. doi: 10.1038/srep01359.

Salim, H. M. W., Ring, K. L. and Cavalcanti, A. R. O. 2008. Patterns of codon usage in two ciliates that reassign the genetic code: *Tetrahymena thermophila* and *Paramecium tetraurelia*. *Protist*., 159: 283-298.

Schultz, D. W. and Yarus, M. 1994. Transfer RNA mutation and the malleability of the genetic code. *J. Mol. Biol.,* 235: 1377–1380.

Schultz, D. W. and Yarus, M. 1996. On malleability in the genetic code. *J. Mol. Evol.,* 42: 597–601.

Söll, D. and RajBhandary, U. L. 2006. The genetic code – Thawing the 'Frozen Accident'. *J. Biosci*., 31: 459-463.

Sussman, J. K., Simons, E. L. and Simons, R. W. 1996. *Escherichia coli* translation initiation factor 3 discriminates the initiation codon *in vivo*. *Mol. Microbiol*., 21: 347–360.

Su, D., Lieberman, A., Lang, B. F., Simonovic, M., Söll, D. and Ling, J. 2011. An unusual $tRNA^{Thr}$ derived from $tRNA^{His}$ reassigns in yeast mitochondria the CUN codons to threonine. *Nucleic Acids Research*, 39: 4866-4874.

Suzuki, T., Ueda, T. and Watanabe, K. 1997. The 'polysemous' codon–a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. *EMBO J.,* 16: 1122–1134.

Swire, J., Judson, O. P., and Burt, A. 2005. Mitochondrial genetic codes evolve to match amino acid requirements of proteins. *J.Mol.Evol*., 60: 128-139.

Tourancheau, A. B., Tsao, N., Klobutcher L. A., Pearlman, R. E. and Adoutte, A. 1995. Genetic code deviations in the ciliates: evidence for multiple and independent events. *EMBO J*., 14: 3262-3267.