July 16-21, 2017 at UC San Diego, CA, USA

The total number of possible genetic codes with 64 triplets, 20 amino acids, and one stop signal

Marco V. José[1] and Eberto R. Morgado[2]

[1]*Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, México CDMX. 04510, México*
[2]*Facultad de Matemática, Física y Computación, Universidad Central "Marta Abreu" de Las Villas, Santa Clara, Cuba*

* e-mail: marcojose@biomedicas.unam.mx

Abstract: The question about the existence of different genetic codes in other regions of the universe is open, as open is the question of the existence of life in other planets of this or other galaxies. If we assume that the four RNA nucleotides C, U, A, G and 20 natural amino acids are the same as in the current standard code, a great number of possible different codes may exist. In this work, we count all the possible genetic codes with some properties in common with the current standard code. In all cases, we assume that there are the 4 RNA nucleotide bases C, U, A, G and 20 amino acids, plus a symbol s with the role of stop signal. We present 3 cases: 1. The subsets A_i of elements of the set $A_S = A \cup \{S\}$, where A denotes the set of 20 amino acids, which have the same number i of coding triplets, being $i \in \{1,2,3,4,5,6\}$, are the same of the current standard code. There are five different kinds of amino acids, according to the numbers of their coding triplets. It is well known that, for every amino acid, and even for the stop signal, the number of coding triplets is equal to 1, 2, 3, 4, or 6. 2. We assume that the subsets A_i of i-coded elements have the same cardinalities of those of the current standard code, that is, the numbers 2, 9, 2, 5 and 3, which conform a partition of the number $21 = 2+9+2+5+3$. 3. In the more general case, we do not assume any restriction about the number of coding triplets for every element of the set A_S. Then, the total number of possible codes will be the total number of surjective functions $F: NNN \to A_S$, from the set NNN, of the 64 triplets, onto the set A_S of the 20 amino acids plus the stop signal. The standard genetic code is, in essence, a surjective function $F: NNN \to A_S$. The fact that this function F is not injective means that the code is degenerated. The calculated numbers permit to calculate the probabilities of having different genetic codes in our universe. Given the recent discovery of several exo-planets which may show conditions for the existence of life, we discuss the present results in terms of how universal may be the universal genetic code.