How Universal is the Universal Genetic Code?
A Question of ExtraTerrestrial Origins.

Marco V. José, Ph.D.[1,2], Eberto R. Morgado, Ph.D.[3], Tzipe Govezensky, Ph.D.[1], Isui Aguilar, Ph.D.[1]

1Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Apartado Postal 70228, Ciudad Universitaria, CP 04510 México DF, México. 2Centro Internacional de Ciencias A. C, Campus UNAM-UAEM, Col. Chamilpa, CP 62210, Cuernavaca, Morelos, México. 3Facultad de Matemática, Física y Computación, Universidad Central "Marta Abreu" de Las Villas, Santa Clara, Cuba.

Abstract

There are evolutionary mechanisms that tend to preserve inherent properties of life. The Universal Genetic Code (UGC) is the best example of a biological system that is considered to be the least capable of evolving, though capable of undergoing dramatic organizational change and increasing in size. Herein, recent work on modeling the evolution of the UGC, assuming a primeval genetic code, are briefly reviewed and discussed. We also discuss the possibility that this primitive code has an extraterrestrial origin, presumably rioorganisms which achieved a primitive form of life long before the Earth was formed. The UGC conserves vestiges of earlier codes, to wit, extended RNA codes type I and II which may have been intermediate states between the primeval preUGD and the UGC. The primeval, the Extended RNA codes, and the UGC show the property of scale invariance in the distance series of their codons. This scale invariance is a clear manifestation of a complex evolving system which reacts to the environment. In this sense, the evolution of the genetic code may be an example of a Lamarckian process. If extraterrestrial life is to be considered as a cosmic phenomenon, non-Darwinian evolution and other types of genetic codes, different from the Earthly UGC, remain a likely possibility.

Keywords: Astrobiology, panspermia, Universal Genetic Code; Primeval Genetic Code; Evolution of the genetic code; Complex systems; Extended RNA codes; Non- Darwinian evolution

1. The Universal Genetic Code

Evolution is generally understood as a change in the frequency of alleles in a given population (Wilson & Bossert, 1971) or as descent with modification (Darwin, 1859). There is no generally accepted definition of life but there is a consensus, even among those who reject Darwinism (e.g., Joseph 2000), that life and its evolution can best be understood in the context of genomic evolution and the genetic code, i.e. genetic material capable of replication, change, and heredity.

A fundamental feature of all life forms existing on Earth is that, with several minor exceptions, they share the same genetic code, i.e., it is *nearly* universal. The only exceptions to this universality are few and far between and exceedingly minor and appear to be limited to a few species, e.g. mitochondrial, *Euplotes*, some ciliate protozoans,

Tetrahymena). However, even these species are believed to have evolved from the same last common ancestor as those species whose genome functions completely by the UGC (Osawa et al., 1992). Most of the variations lie in the unicoded amino acids (Met and Trp) and in the stop codons. In these species, twenty natural amino acids are encoded, with the notable exception of selenocysteine and pyrrolysine, for which subsets of organisms have evolved special coding schemes (Ambrogelly et al., 2007).

This universality led Crick to propose the frozen accident hypothesis (Crick, 1968), i.e. the code does not change. The universality of the Universal Genetic Code (UGC) suggests to most but not all scientists, that all of life on Earth descended from a Last Universal Common Ancestor, whose own ancestry leads a first life form. The consensus view is life must have arisen just once on Earth after a fortuitous series of chance occurrences, and this explains the UGC; otherwise, it would be expected that there might be a variety of genetic codes or even life forms without genes.

Likewise, the consensus is that this first life form replicated and duplicated its genes and passed on its genome to its descendants. According to Crick (1968), the genetic code remained universal because any change would be lethal, or would have been very strongly selected against and extinguished.

The generally accepted view is that the first living entities were systems capable of replicating and whose self-sustaining properties depended on the availability of organic molecules already present in the primitive environment. Once a general mechanism existed for self-replication, allowing the introduction of variation and the ability to replicate those variants, Darwinian evolution began to operate. This marked the beginning of the evolution of of life.

2. Horizonatal Gene Transfer

Because the genetic code is universal, this enables even completely different species to shares genes which can become incorporated into the genomes of even alien species. Vetsigian and coworkers (2006) for example, have recently modeled the universality of the UGC by considering early communal evolution of cell-like entities mediated by horizontal gene transfer (HGT).

Some believe that gene sharing contributed to the fashioning of the first eukaryotes and has promoted species diversity and evolution (Joseph, 2009a). Woese, (2004), for example, has proposed that early in the Earth's history, primitive proto-bacteria, proto-archaea and proto-eukaryotes may have lived together and shared genes via HGT. "Eventually this collection of eclectic and changeable cells coalesced into the three basic domains known today" (Woese, 2004).

It has also been proposed that the first Earthly unicellular eukaryotes were fashioned when genes from archae, bacteria, and viruses combined thereby inducing eukaryogenesis and giving rise to the eukaryote genome (Joseph, 2009a). A number of investigators now believe HGT has played a significant role in the evolution of species and the genome, and has contributed to evolutionary innovation.

3. Genes and the Environment

The genome, its genes, and the genetic code play a dual role: acting not only as mechanisms for encoding amino acid sequences and expressing and building proteins, but also an innovation-sharing protocol. However, there are yet other important factors that only a few investigators have recognized: the environment acts on gene selection, such that evolution of species and the genome involves a complex gene-environmental interaction (Joseph 2000, 2009a;

Vetsigian et al., 2006). Based on their analysis utilizing different models, Vetsigian and coworkers (2006) have concluded that the "evolution of the genetic code, translation, and cellular organization itself follows a dynamic whose mode is, if anything, Lamarckian."

The great diversity of living beings in the history of the biosphere has not been halted by a frozen UGC. The inherent structure of the frozen UGC, in concert with environmental influences, has unleashed life from determinism.

4. The RNA World

All life that is known to exist on Earth today and all life for which there is evidence in the geological record seems to be of the same form — one based on DNA genomes and protein enzymes. Yet there are strong reasons to conclude that DNA- and proteinbased life was preceded by a simpler life form based primarily on RNA. This earlier era is referred to as the "RNA World", during which the genetic information resided in the sequence of RNA molecules and the phenotype derived from the catalytic properties of RNA (Gilbert, 1986).

The RNA molecule is ubiquitous in all current biological systems, especially with regard to the most fundamental and highly conserved cellular processes. It is involved as a primer in DNA replication, a messenger that carries genetic information to the translation machinery, and a catalyst that lies at the heart of the ribosome. RNA instructs the processing of precursor messenger RNAs during splicing and editing, and mediates numerous other transactions of RNA and proteins in the cell (Joyce, 1991).

Catalytic RNAs (ribozymes) assist in RNA processing events and the replication of viral genomes. Individual nucleotides serve as important signaling molecules and their coenzyme derivatives participate in most of the reactions of central metabolism.

The code is implemented via the tRNAs that bind each codon with its anticodon. The anticodon is located in the middle of the tRNA cloverleaf, at the maximum possible distance where the cognate amino acid will be attached. Because of this separation, tRNA molecules cannot self aminoacylate; instead, there are 20 amino acid-specific aminoacyl-tRNA synthetases that perform this function. These molecules define the genetic code, by linking the specific amino acids and tRNAs with the corresponding anticodons (Rodin & Rodin, 2008).

RNA has several features that make it suitable as the basis for a simple Darwinian system: it contains only four different subunits with very similar chemical properties, its subunits polymerize readily when activated and bound to a complementary template, it is a polyanion that is readily soluble in water almost irrespective of sequence, it forms simple secondary structures that are highly tolerant of sequence variation, and it can adopt entirely different structures following the acquisition of a few critical mutations (Fontana & Schuster, 1998; Schultes & Bartel, 2000).

These same features make it less sophisticated compared to its DNA and protein successors. The lower reactivity but greater stability of DNA makes it a better choice for the genetic material, whereas the greater chemical diversity of the subunits of proteins, including anionic, cationic and hydrophobic groups, makes protein a better choice as the basis for catalytic function. However, those more sophisticated molecules could not have arisen without the foundation that had been laid by RNA.

5. An ExtraTerrestrial Primeval Code?

There are substantial gaps in scientific understanding concerning how the RNA world arose, the degree of metabolic complexity that it attained, and the way that it led to DNA genomes and protein enzymes. In fact, some scientists have called the RNA world a myth (Joseph 2009b) and have pointed to the fact that viruses, with their RNA genome, still require the DNA of a host to replicate. However, viruses with a DNA-genome also require the DNA of a host to replicate.

If we assume that life had an origin beginning with some type of proto-cell which could self-replicate, and that RNA and DNA replicative mechanisms are the result of a progressive evolution over billions of years of time, then it is possible that this primitive code has an extraterrestrial origin. Indeed, statistical probability argues against the odds of random events creating fully formed life on Earth from scratch. Therefore, the steps leading to an RNA world began before the Earth was formed, perhaps even 10 billion years ago (Sharov 2010). We propose that the beginning of what would become universal genetic code may have begun to evolve in extra-terrestrial riboorganisms which achieved a primitive form of life long before the Earth was formed. These riboorganisms could then have been dispersed throughout our galaxy through mechanisms of panspermia (cf. Arrhenius 1908/2009; Burchell 2010; Joseph 2009b; Rampelotto 2009).

The reign of the RNA world on Earth probably began no more than about 4.2 billion years ago and ended no less than about 3.6 billion years ago (Joyce, 2002). Eigen and coworkers (1989) revealed kinship relations by alignments of tRNA sequences and they concluded that the genetic code is not older than, but almost as old as our planet. This finding ruled out the possible cosmic origin of the Earthly code. They also stated that "had the code been much older- and this would be possible only in case of extraterrestrial origin-those changes that clearly can be identified as phylogenetic divergence would previously have become randomized to a large extent". It is a general belief that life originated on Earth, yet there is no direct experimental evidence about when and where life actually came about. Therefore, the understanding of the UGC could lead us to propose other types of plausible genetic codes if extraterrestrial life is to be invoked.

Recently, two genetic codes from which the primeval RNA code could have originated the UGC were derived (José et al., 2007). The primeval RNA code consists of 16 codons that specify 8 amino acids (then this code shows a slight degeneration). The Extended RNA code type I, consists of all codons of the type RNY (purine-any base-pyrimidine) plus codons obtained by considering the RNA code but in the second (NYR type) and third (YRN type) reading frames. The extended RNA code type II, comprises all codons of the type RNY plus codons that arise from transversions of the RNA code in the first (YNY type) and third (RNR) nucleotide bases.

Then, by allowing frame-reading mistranslations we arrive at 48 codons that specify 17 amino acids and the three stop codons. If transversions in the 1st or 3rd nucleotide bases of the RNY pattern are permitted, then there are also 48 codons that encode for 18 amino acids but no stop codons.

The codons in each of the subsets of both Extended RNA codes were represented by four-dimensional symmetrical hypercubes (José et al., 2007), whose union comprised precisely the already known six-dimensional hypercube of the UGC of 64 triplets (Jiménez-Montaño, et al., 1996). Evolutionary analysis of UGC based upon novel three-dimensional algebraic models, dubbed Genetic Hotels, lead more clearly to the same conclusions (José et al., submitted).

The set of the 64 triplets has been represented as a GF(4)-vector space, being GF(4) the Galois field of four elements. In this way, the set is inserted in the 3-dimensional Euclidean vector space and it is envisaged as a multicube, that is,

a cube with inner points and points in its faces and its edges. With this geometrical approach it happens that most of the amino acids have sets of coding triplets that are unions of vertical edges of the multicube.

If we constrain ourselves to the mitochondrial genetic code we find that there are neither unicoded amino acids nor three coded, but every amino acid has a set of coding triplets with an even, (2,4 or 6) number of elements. It can be stated the following, Theorem: In the 3-dimensional mitochondrial genetic code (Genetic Hotel), every amino acid has a set of coding triplets, which is a union of vertical edges.

The Extended RNA codes type I and II, represent two plausible evolutionary paths of the different prokaryote genomes since the RNA World to current prokaryote genomes. In the context of the frozen concept, it was concluded that considering the symmetries of both Extended RNA codes, the primeval RNY code was already frozen and that it evolved like a replicating and growing icicle (José et al., 2007).

Arguably the life forms that probably obeyed the Extended RNA code type I and II were intermediate between the ribo-organisms of the RNA World and the cenancestor of Bacteria, Archea, and Eucarya. These results support the notion that evolution did not erase all vestiges typical of the RNA world in today's prokaryote genomes, not only in terms of an enrichment of RNY codons, but also revealing the existence of an underlying ancient fractal structure. Indeed, the scaling properties of the distance series of some codons from the RNA code and most codons from both Extended RNA codes turned out to be identical or very close to the scaling properties of codons of the UGC (José et al., 2009).

To test for the robustness of these results, it was shown, via computer simulation experiments, that random mutations of current genomes, at the rates of 10-10 per site per year during three billions of years, were not enough for destroying the observed patterns. Therefore, most current prokaryotes still contain relics of the primeval RNA world and both Extended RNA codes may well represent two plausible evolutionary paths between the RNA code and the current UGC.

Since the magnitudes of the scaling exponent of the distance series of these codons are unaltered throughout evolutionary time, we can say that they show critical scale invariance according to the renormalization group theory (José et al., 2009). In other words, these scaling and statistical properties of the distance series of the codons of a given prokaryote did not change during the evolution of the UGC.

Thus, several scaling and statistical features of the RNA and the extended RNA type I and II codes are still detectable in current prokaryote chromosomes since they show a critical state, with the interesting exception, to some extent, of some Archeas. Therefore, genomes are systems that are constantly under a critical state and they may show universal properties of scale invariance.

This scale invariance of certain sequences of RNA/DNA is the hallmark of complex systems. Complexity is not explained by Darwinian evolution which is concerned with variability upon which natural selection operates.

However, there are several self-organizing biological systems that lead to highly ordered structures that must operate, in addition to natural selection.

Unfortunately, complexity models have been dismissed because invocations to spontaneous generation appear to be lurking behind appeals to undefined -emergent properties, or -selforganizing principles, that are used as the basis for

what many life scientists see as grand, sweeping generalizations with little, if any, relationship to actual biological phenomena (Lazcano, 2007).

Despite the long time that has passed there seems to remain original vestiges surviving in current DNA prokaryote sequences. Regardless of the details of RNA/DNA sequences, an emergent property arises: a simple scaling exponent of the distance series of a given codon has been the same during the evolution of the UGC in a given organism.

Therefore, we believe the evidence indicates that this primitive, pre-RNA world code has an extraterrestrial origin, presumably riboorganisms which achieved a primitive form of life long before the Earth was formed.

6. Conclusions

In summary, there is a strong support of the attractive and appealing idea of fossilized remnants of the primeval and the Extended RNA codons in today existing genes. We believed these may have their origins in primitive extraterrestrial riboorganisms.

So far all we know is terrestrial biology. Yet if we are to consider life as a cosmic phenomenon, we must be open to the possibility that life may be everywhere, and that the "seeds of life" swarm throughout the cosmos (Arrhenius 1908/2009; Joseph 2009b) such that life on Earth is just a sample of life's possibilities, with all of life evolving from the same universal genetic code which has cosmic origins.

According to Sharov (2010) and Joseph (2000, 2009a) if life on Earth came from an ancient cosmic source, and if these life forms were spread among the planets according to the mechanisms of panspermia, then extraterrestrial life should be genetically identical to terrestrial life and extraterrestrial prokaryotes may even be governed by the same universal genetic code, which would indicate the shared cosmic ancestry of all life in the solar system.

This scenario is in agreement with the theory of panspermia originally proposed by Arrhenius (1908/2009). According to Arrhenius, life-giving seeds are drifting about in space propelled by the tiny but cumulative pressure of starlight. The seeds encounter the planets and fill their surfaces with life as soon as the necessary conditions for the existence of organic beings are established. Yet other proponents of panspermia have proposed that life journeys from world to world via comets, asteroids, and planetary debris, exchanging DNA via horizontal gene transfer, and this is how life arrived on Earth (Joseph 2000, 2009a,b). According to this view, since life on Earth came from an extraterrestrial source, and since all Earthly life evolved from the first life on Earth whose genetic ancestry is extraterrestrial, then life should evolve in a similar fashion on other Earth-like planets, since all life is derived from the same ancestral extraterrestrial DNA. This non-Darwinian perspective views evolution as a form of metamorphosis and the result of complex gene-environmental interactions (Joseph 2000, 2009a).

By contrast, if life arose independently via mechanisms related to abiogenesis, then we have to consider the possibility that these life forms have different types of genetic codes (Sharov 2010). Therefore, life on other worlds may have *evolved* according to principles which are alien to Darwinian concepts of evolution. However, again, even on Earth we see evidence of non-Darwinian patterns of evolution, possibly Lamarckian. Therefore, regardless of how life originates on these other worlds we must consider the possibility of other types of genetic codes perhaps evolving to complexity in a Lamarckian-like fashion. This may be the case even if the genomes of life on other planets can be traced to primitive extraterrestrial riboorganisms which first began replicating 10 billion years ago.

As to Earth, it is commonly assumed that life came from non-life, via abiogenesis, and that the first cells on Earth had to be pieced together by a series of random accidents involving chemical substances already at hand in the Earth's waters mixing with organic material which originated in comets, meteors and asteroids (Menor-Salván 2009; Sidharth 2009). However, the statistical probability of such an event occurring even after 4 billion years years of random mixing, can be seen as vanishingly small. Give that complex life appeared anywhere from 500 million to a billion years after the Earth was fashioned, to achieve, in such a short span of time, essentially all at once, a flawless code for the complex information system of DNA and RNA, which would later become the UGC for all Earth's creatures, has seemed so improbable that Francis Crick and Leslie Orgel regarded it as outright impossible. Instead Crick and Orgel (1973) seriously proposed that prokaryotes must have been planted here via "directed panspermia" by highly advanced extraterrestrials over 4 billion years ago; bucking the problem of life's origin to someone else living someplace else in the cosmos.

According to our work (José et al., 2009), the evolution of the primeval genetic code occurred on Earth, but did not originate on Earth. If this primitive code has an extraterrestrial origin, the original seeds, presumably riboorganisms, could have evolved towards different genetic codes in different planets and helped to jump start the RNA world which led to complex life on this planet. If life is to be found somewhere else in the universe, and even if their ancestry can be traced back to these same riboorganisms, the rules of assignments of codons to aminoacids may not necessarily be the same and the aminoacids may be even chemically different to those found in known terrestrial life. Different environments, and different evolutionary paths on different worlds could result in completely different genetic codes and patterns of evolution.

In conclusion, the UGC is certainly ubiquitous in Earth and what we would expect to find in living beings on other planets is a biological universal function: a genetic coding system. However, from the perspective of cosmological biology, these genetic codes may not be universal but unique to the life forms on the planets where they evolved.

References

Ambrogelly A, Palioura S, Soll D. (2007). Natural expansion of the genetic code. Nat Chem Biol. 3(1):29-35.

Arrhenius S. (1908). Worlds in the Making. The Evolution of the Universe. Harper & Brothers, New York and London.

Arrhenius S. (2009). The Spreading of Life Throughout the Universe. Journal of Cosmology, 2009, 1, 91-99.

Burchell, M.J. 2010. Why Do Some People Reject Panspermia? Journal of Cosmology, 5, 828-832.

Crick F. H. C. (1968). The origin of the genetic code. J. Mol. Biol. 38: 367-379.

Crick, F. 1981. Life Itself. Its Origin and Nature. Simon & Schuster, New York.

Crick, F. H. C., and Orgel, L. E. (1973). Directed Panspermia," Icarus, 19, 341.

Darwin, C. (1859). On the Origin of Species by Means of Natural Selection. London: John Murray.

Eigen, M., Lindemann, B. F., Tietze, M., Winkler-Oswatitsch, R., Dress A., Haeseler A. (1989) How old is the genetic code? Statistical geometry of tRNA provides an answer. Science 244: 673-679.

Fontana, W., Schuster, P. (1998). Continuity in evolution: on the nature of transitions. Science 280, 1451–1455.

Gilbert, W. (1986). The RNA world. Nature 319, 618.

Jiménez-Montaño, M.A., de la Mora-Basañez, C.R., Pöschel, T. (1996). The hypercube structure of the genetic code explains conservative and non-conservative amino acid substitutions in vivo and in vitro. Biosystems 39, 117–125.

José, M. V., Morgado, E. R., Govezensky, T. (2007). An extended RNA code and its relationship to the standard genetic code: an algebraic and geometrical approach. Bull. Math. Biol. 69: 215-243.

José, M. V., Govezensky, T., García, J. A., Bobadilla, J.R. (2009). On the evolution of the standard genetic code: vestiges of scale invariance from the RNA World in current prokaryote genomes. Plos ONE. February 2, vol. 4, e4340.

José, M. V., Morgado, E. R., Govezensky, T. Genetic Hotels for the Standard Genetic Code: Evolutionary Analysis based upon Novel Three-dimensional Algebraic Models. Submitted: Bull. Math. Biol. Joseph, R. (2000). Astrobiolgy, the Origin of Life and the Death of Darwinism. University Press. California.

Joseph, R. (2009a). The evolution of life from other planets. Journal of Cosmology, 1, 100-200.

Joseph, R. (2009b). Life on Earth came from other planets. Journal of Cosmology, 1, 1-56.

Joyce, G. F. (1991). The rise and fall of the RNA world. New Biol. 3, 399–407.

Joyce G. F. (2002). The antiquity of RNA-based evolution. Nature 418: 214-221.

Lazcano, A. (2007). Prebiotic evolution and the origin of life: is a system-level understanding feasible? In Systems Biology. Volume I: Genomics., Eds. I. Rigoutsos and G. Stephanopoulos, Oxford University Press, New York, 2007, p. 57–78.

Menor-Salván, C. (2009). Abiogenesis: Life May Have Begun on Earth. Journal of Cosmology, 2009, 1, 71-72.

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

Rampelotto, P. H. 2009. Are We Descendants of Extraterrestrials? Journal of Cosmology, 2009, 1, 86-88.

Rodin, S. N., Rodin A. S. (2008). On the origin of the genetic code: signatures of its primordial complementarity in tRNAs and aminoacyl-tRNA synthetases. Heredity 100: 341-355.

Schultes, E. A., Bartel, D. P. (2000). One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science 289, 448–452.

Sidharth, B. G. (2009). In Defense of Abiogenesis. Journal of Cosmology, 2009, 1, 73-75.

Sharov, A. A. 2010.Genetic Gradualism and the ExtraTerrestrial Origin of Life. Journal of Cosmology, 5, 833-842.

Vetsigian, K., Woese, C., Goldenfeld, N. (2006). Collective evolution and the genetic code. Proc. Natl. Acad. Sci. USA 103: 10696-10701.

Wilson E. O., Bossert, W. H. (1971). A Primer of Population Biology. Sinauer Associates, INC. Publishers, Sunderland, Massachusetts.

Woese, C. (1968). The Genetic Code. Harper & Row.