BioSystems xxx (xxxx) xxx

Contents lists available at ScienceDirect

BioSystems

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

Editorial

Overview of the third special issue in code biology

1. Introduction

This third special issue in Code Biology is a collection of highly different papers and their differences have two main causes. The first, the most obvious, is that Code Biology is the study of all codes that exist in living systems and the diversity of the papers is a direct consequence of the diversity of the codes. The second source of diversity is the existence of different theories. More precisely, the *original theory* that gave origin to Code Biology has been followed by a number of *extended theories* that now coexist with the original one. In Code Biology, in other words, there is *pluralism* but there has also been a *beginning*, and it is important to be clear about this starting point. The original theory of Code Biology is characterized by ideas that make it different from four major theoretical frameworks:

[1] The original theory of Code Biology is different from the *Modern Synthesis* for two reasons. The first is the idea that evolution took place by natural selection and by natural conventions and these mechanisms are fundamentally different because natural selection is based on copying and natural conventions are based on coding. The second is the idea that the cell is not a biological computer made of genotype and phenotype but a trinity of genotype, phenotype and ribotype, where the ribotype is the ribonucleoprotein system of the cell that functions as the codemaker of the genetic code (Barbieri 1981, 1985, 2003).

[2] The original theory of Code Biology maintains that the fundamental process of life is not autopoiesis but codepoiesis (Barbieri 2012). Autopoiesis requires biological specificity and specificity comes from the genetic code, so the ancestral systems that came before that code could not have been autopoietic systems. Those ancestral systems, on the other hand, were engaged in the evolution of the genetic code and were therefore codepoietic systems. Autopoiesis, furthermore, is most evident in bacteria and bacteria have not increased their complexity and have not evolved new codes for billions of years after their appearance on Earth. It was the eukaryotes that became increasingly complex and that evolved new codes, which suggests a deep link between codes and complexity, and in particular between the origin of new codes and the origin of the great novelties of macroevolution (Barbieri 2015, 2016, 2017, 2020). Codepoiesis, on the other hand, is necessarily implemented by mechanisms, and according to the original theory of Code Biology the major mechanism that fuelled the evolution of the genetic code was the process of *ambiguity reduction* (Barbieri 2019a).

[3] The original theory of Code Biology is different from *Biosemiotics* because it claims that the Peircean processes of *interpretation* and *abduction* take place in the brain but not in the cell (Barbieri 2014, 2018).

[4] The original theory of Code Biology is different from the *Relational Biology* of Robert Rosen because it assumes that the process of *anticipation* takes place in the brain but not in the cell (Barbieri 2019b).

There are, in conclusion, four key ideas in the original theory of Code Biology:

[a] Evolution took place by natural selection and by natural conventions.
[b] The cell is a trinity of genotype, phenotype and ribotype.
[c] The fundamental process of life is codepoiesis, not autopoiesis.
[d] Ambiguity reduction was the major evolutionary mechanism of the genetic code.

The extended theories of Code Biology differ from the original theory either because they introduce new concepts or because they reformulate some of the original concepts.

[1] The first extended theory appeared when Stefan Kühn and Jan-Hendrik Hofmeyr (2014) proposed an *extended definition of code*, a definition where signs and meanings can be not only molecules but also biological processes. More precisely, Kühn and Hofmeyr showed that the histone code is a mapping where the signs are the marks produced on histones by acetylation or methylation processes and their meanings are the activation or the repression of particular genes.

[2] A second extended theory of Code Biology has been proposed in this issue by Julie Heng and Henry Heng with the idea that the adaptors of a biological code can be "*information flows*". More precisely, Heng and Heng point out that in addition to the codes that produce the components of a system there are also codes that organize those components into a working whole. The code that is used to make bricks, for example, is different from the code that is used to construct a building from those bricks. The genetic code is a code that makes bricks, i.e., proteins, but in order to arrange proteins into a living system we need an architectural code that Heng and Heng call "*karyotype code*".

[3] A third extended theory is presented in this issue by Omar Paredes and colleagues on the grounds that the original theory of Code Biology "*raises the illusion that information has only an*

https://doi.org/10.1016/j.biosystems.2021.104553

Available online 27 September 2021

> *upward direction … whereas the current overview of cellular dynamics … illustrates that information flows freely upward and downward*". In order to overcome this limitation, the authors propose "*a novel category of organic codes, the metacode*", which is defined as "*an informational structure that handles the continuum of the information flow in biological systems*".

The extended theories, in short, are a reality and their existence is a testimony that there is genuine pluralism in Code Biology. The goal of this special issue, on the other hand, is to give a bird's-eye view of the present status of Code Biology and to this purpose it has been divided into four parts, each of which is going to be illustrated in the rest of this editorial with brief presentations of its papers.

2. PART 1 – Codes and evolution

(Igamberdiev, 2021; Kun, 2021; Heng and Heng, 2021; Dieci, 2021; Coca et al., 2021).

1 – Abir Igamberdiev

The *drawbridge of nature: Evolutionary complexification as a generation and novel interpretation of coding Systems*.

The author underlines that Lamarck was the first to distinguish between two types of evolution, one that changes the morphology of an organism without changing its complexity and a second one that leads to an overall increase in complexity. In 1809 Lamarck called this complexification "*Le pouvoir de la vie"*, in 1911 Henry Bergson called it "*creative evolution*", and in 1932 Osip Mandelstam called it "*the drawbridge of nature*", the name that gives the title to this paper. An evolution that takes place with an increase in complexity has also been referred to as "*progressive evolution*" and its causes have been one of the most controversial issues of biology.

The original theory of Code Biology reached the conclusion that absolute novelties can only be created by new codes and proposed therefore that progressive evolution is based on the generation of new codes, a process that was referred to as *codepoiesis*. The author of this paper agrees that codepoiesis is a mechanism of complexification but then he takes a step further. The goal of his paper is to define the basic principles of codepoiesis, and to this purpose evolutionary complexification is described as a metasystem transition that "*expands the system by incorporating externality*". In this process, the values of a system are assigned to previously unproven statements, which shows that during evolution living systems "*continuously realize the proof of Gödel's theorem*".

The idea that a code is generated by incorporating externality was not present in the original theory of Code Biology and represents therefore a new development in this field. In the original theory an absolute novelty is something that did not exist before, neither inside nor outside the system, but the idea that the environment can act as an *invisible hand* in the evolution of new codes is an interesting possibility and should be the object of further investigations.

2 – Ádám Kun

The Major Evolutionary Transitions and Codes of Life.

In '*The Major Transitions in Evolution*' John Maynard Smith and Eörs Szathmáry (1995) declare, at the very beginning of the book, that they are "*committed"* to the view that natural selection is the sole mechanism of evolution. But where does natural selection come from? The genes are transmitted from one generation to the next by molecular copying, but when copying is repeated indefinitely mistakes become inevitable and not all changes can survive so a selection is bound to take place. Molecular copying, in short, leads to heredity, and the indefinite repetition of molecular copying in a world of limited resources leads to *natural selection*. This is how natural selection came into existence. Molecular copying started it and molecular copying has perpetuated it ever since. Which means that *natural selection would be the sole mechanism of evolution if molecular copying is the sole basic mechanism of life*.

The discovery of the genetic code, however, has proved that there are *two* molecular mechanisms at the basis of life, the *copying* of the genes and the *coding* of proteins. Life, in other words, is not based on copying alone. It is based on copying and coding, and these two molecular mechanisms give origin to two distinct mechanisms of evolution because an evolutionary mechanism is the long term result of a fundamental molecular mechanism. More precisely, the existence of copying and coding at the molecular level means that there are two distinct types of evolutionary change: *evolution by natural selection*, based on copying, and *evolution by natural conventions*, based on coding (Barbieri 1985, 2003).

According to Maynard Smith and Szathmáry, in short, evolution took place by natural selection only, whereas according to Barbieri evolution took place by natural selection and by natural conventions. This is the key issue of the Major Transitions: were they caused by one or by two mechanisms?

The crucial point, here, is that without a code a Major Transition simply would not take place. Without the genetic code, for example, biological specificity would not exist and the Major Transition from the RNA world to the protein world would not have been possible. The Major Transitions, in other words, are the historical events that brought *absolute novelties* into existence, and absolute novelties cannot exist without new codes.

The author of this paper comes from a school which claims that all Major Transitions were caused by natural selection only, but organic codes are experimental realities and he has therefore the problem of finding out what was their role in the great events of macroevolution. His present paper is the beginning of a meditation on this fundamental problem of life.

3 – Julie Heng and Henry Heng

Karyotype coding: the creation and maintenance of system information for complexity and biodiversity.

The authors of this paper propose that there are two very different types of codes in life: the codes for producing the components of a system ("*parts codes*") and the codes for assembling those components into a working whole ("*system codes*"). They point out that the organic codes which have been discovered so far (genetic code, histone code, splicing codes, signal transduction codes, tubulin code, lamin code, Hox code, etc.) are all *parts codes* and yet there must also be a *system code* in every cell in order to ensure that the same organization is reconstructed in its descendants. This system code has been called by the authors "*karyotype code*" because the karyotype is the highest level of genomic topological information.

The difference between the karyotype code and the other codes comes from their *adaptors*. Any organic code is a set of rules that establish a mapping between two independent worlds by means of adaptors and is many cases adaptors are organic molecules but this is not necessarily true in all cases. In the karyotype code, for example, the adaptors establish a mapping between upstream and downstream flows of information and this can only be achieved by populations of molecules not by individual molecules. The authors conclude therefore that the definition of adaptor must be generalized and this is a genuine extension of the original theory of Code Biology.

In order to underline the importance of the karyotype code, the authors recall the Human Genome Project, a massive enterprise that started in the 1990s with the great expectation of disclosing what makes us human. Despite costing enormous resources and generating massive data sets, little was revealed about the actual working of the genome, and this, according to the authors, is because the genetic code tells us how to make the bricks of a body (the proteins) but we still do not have the code that assembles those bricks into a house.

4 – Giorgio Dieci

Removing quote marks from the RNA polymerase II CTD 'code'.

The expression of genetic information begins with the transcription of a gene into a messenger-RNA, an operation that is performed by enzymes called *RNA-polymerases*. A gene is a sequence of DNA-nucleotides, whereas a messenger-RNA is a sequence of RNA-nucleotides, so the function of the RNA-polymerases is to

scan a gene, recognize its DNA-nucleotides and associate to each of them a specific RNA-nucleotide. This operation is performed automatically on all genes of all cells in all living systems, and in principle it could be realized by a single universal RNA-polymerase, but this is precisely what does *not* happen. Bacteria have RNA-polymerases that are different from the eukaryotic ones, and the eukaryotes have three different types of RNA-polymerases: one for the genes that code for ribosomal RNAs, a second for the genes that code for proteins, and a third one for the genes that code for transfer-RNAs and for untranslated RNAs.

The vast majority of the transcription operations is carried out by the second polymerase, but even in this case the final results can be subject to variations. This is because the end of the polymerase contains a string of seven peptides that is repeated a variable number of times (26 in yeast, 52 in chordates, etc.) and the amino acids of these peptides can be modified *after* translation, a process that leaves a '*quote mark*' on the polymerase.

The key point is that these quote marks are recognized by different proteins and the end result is that the same gene can be transcribed into different messenger-RNAs. This variability, on the other hand, must take place according to rules and these rules have been referred to as the "*RNA-polymerase code*" (technically the "*RNA polymerase II CTD code*"). The problem with this code is that, for a variety of reasons, it has always been regarded as a metaphorical entity, not as a real code, but in this paper the author demonstrates that it has all the characteristics that define a real organic code. We have therefore a new set of rules in protein synthesis and a new code in molecular biology.

5 – Juan Coca, Hasier Eraña and Joaquín Castilla

Biosemiotics comprehension of PrP code and prion disease.

The discovery of prions and of their devastating pathological consequences has produced a tsunami in biology because it has raised the possibility of a protein-based transmission of information. Prions have been described by Stanley Prusiner (1982) as the sole responsible for various neurodegenerative diseases that are collectively known as *transmissible spongiform encephalopathies*. These pathologies are caused by prion-proteins that can acquire a misfolded structure and transmit this structure to other prion-proteins without any genetic change, which amounts to saying that "*biological information can be encoded in the structure of a protein and does not rely only in genetic variation*".

Countless papers on prions and on their diseases have been published since 1982, but the article that appears in this issue has a unique feature: it is the first to propose that there is a *code* at the heart of the prion mechanism. The authors point out that the proteins which utilize misfolding to regulate their functions have been found in many distant species such as bacteria, fungi, gastropods and mammals, and it is possible therefore that "*all of them share a general common biological code*". Unfortunately, the experimental evidence that we have today tells us that the existence of a prion code is a *possibility*, but not yet a certainty. That possibility, however, is undoubtedly worth exploring and this paper has the merit to open up an entirely new approach to the still largely mysterious universe of the prion proteins.

3. PART 2 – Codes and biology

(Hofmeyr, 2021; Paredes et al., 2021; Petoukhov, 2021; Zámečník, 2021; Faria, 2021; Major, 2021).

6 – Jan-Hendrik Hofmeyr

A biochemically-realisable model of the self-manufacturing cell.

The author underlines that "*in order to grow, reproduce, metabolise, self-maintain, adapt and evolve, living organisms must first and foremost be able to self-manufacture*". This is therefore the hallmark of life, the bedrock that divides living from non-living systems: "*the ability of living organisms to manufacture and individuate themselves autonomously as wholes*". This ability of the cell to fabricate its own components was called *autopoiesis* by Maturana and Varela (1980) and was illustrated by Robert Rosen with a model called *Metabolism-Repair system*, or (M, R)-system, which states that a system is alive if and only if it is "*closed to efficient causation*".

The problem with these proposals is that they have not been translated into real biochemical processes, and it is this obstacle that Hofmeyr has addressed and overcome in the present paper. He has divided the fundamental processes of the cell into three great classes: (1) the enzyme catalysts of covalent metabolic chemistry, (2) the intracellular milieu that drives the supramolecular processes of folding and self-assembly, and (3) the membrane electrolyte transporters that maintain the intracellular milieu. This has allowed him to describe the cell as a *Fabrication-Assembly system*, or (F,A)-system, that, unlike Rosen's (M,R)-system, has a clear biochemical realization.

Hofmeyr maintains that his model of the cell accommodates and gives biochemical substance to concepts such as Robert Rosen's relational biology, Howard Pattee's epistemic cut, Marcello Barbieri's genotype-ribotype-phenotype ontology, and Tibor Gánti's chemoton. Given the historical opposition between these concepts, such a conclusion may be a bit optimistic, but the goal of an ultimate unity is certainly worth pursuing and Hofmeyr's model of the cell does point in that direction.

7 – Omar Paredes, Alejandro Morales, Adriana Mendizabal and Rebeca Romo-Vázqueza

Metacode: one code to rule them all.

The authors of this paper declare that they are going to propose "*a novel category of organic codes, the metacode*", a proposal that "*leads to a new domain in code biology*". And they clearly say why they are doing this. The reason is that the original theory of code biology "*raises the illusion that information has only an upward direction … whereas the current overview of cellular regulatory dynamics … illustrates that information flows freely upward and downward*" which means that organic codes "*must be interlinked rather than just one-way mappings*".

In order to implement this project the authors declare that they must first clear the ground and develop "*a theoretical framework where life is not placed as an extraneous entity detached from the universe's laws*". This is necessary because "*so far the organic codes have been conceptualized as self-enclosed units in their levels containing all that is required to carry out phenomena within that level*", an approach that is "*off-putting, as it reminds ominously of particle science and its quest to complete the particle zoo.*"

The new theoretical framework proposed by the authors is one where *"biological systems are informatic singularities that navigate an energy metaspace, thus building structures that increase the entropy in the universe. Through complexity-noise balance, they stall in niches where organic codes take meaning. Transiting those niches signifies increasing complexity where a wider information flow is transmitted up and downwards through its organizational levels, a feature that organic codes so far do not address."*

The conclusion of this paper is that "*the present code biology framework is updated, calling for a review of organic codes under this new framework*". But is this true? Perhaps one day 'Code Biology' will be replaced by 'Metacode Biology', but for the time being this does not seem an imminent possibility.

8 – Sergey Petoukhov

Algebraic harmony and probabilities in genomes. Long-range coherence in quantum code biology.

This paper is an assembly of four theories that are largely independent and it is possible therefore to examine them separately.

[1] The first theory is a generalization of the second Chargaff's rule which states that the percentages of C and G are equal to the percentages of A and T. This generalization is illustrated by numerous tables and figures and looks like a valuable contribution to biology.

[2] The second theory is the idea that the mechanisms behind genetic information and organic codes come from quantum mechanics and quantum informatics. The author agrees with Jordan that "*life's missing laws are the rules of chance and probability of the quantum world*", a claim that characterizes what has been called "*the quantum approach to life*". The problem here is that nobody has described what these quantum rules of life actually are, so it is not possible to express a scientific opinion about them.

[3] The third theory is the idea the genetic rules are *n*-plets alphabets of DNA whose nucleotide sequences are considered as parallel texts written in interconnected *n*-plets alphabets, a conclusion that is based on the results obtained by the author in his studies of eukaryotic and prokaryotic genomes.

[4] The fourth theory is the idea that many processes of life can be explained by the resonance model proposed by Fröhlich (1970). This model argues that long-range vibrations can induce a resonance between a differentiated cell and its chromosomes and that the chromosome parts that respond to particular frequencies become activated and produce the appropriate proteins. Such a resonance, furthermore, could transport the embryonic cells to their target where they would be further induced into producing the correct proteins through superimposed resonances.

What this paper is proposing, in conclusion, is an approach to biology that gives a central role to the processes of quantum mechanics and quantum informatics and the author summarizes this approach with the statement that **"***Life is a partnership between genes and mathematics*".

9 – Lukáš Zámečník

Causal and non-causal explanations in Code Biology.

This paper acknowledges that there are three different paradigms in modern biology: (1) the chemical paradigm, which states that "*Life is chemistry*", (2) the information paradigm, which states that "*Life is chemistry + information*", and (3) the code paradigm, which states that "*Life is chemistry + information + codes*". Modern biology has been dominated by the debate between the first two paradigms but a solution has never been found because the chemical paradigm has not been able to reduce information to physics and chemistry, whereas the information paradigm has not been able to explain why information is *ontologically* different from the quantities of physics and chemistry.

A way out of this stalemate has been proposed by Code Biology with the idea that information is a fundamental *observable*, i.e., an entity that is essential to describe what we *observe* in nature. The world of electromagnetism, for example, is different from the world of gravitation because its description requires a new observable, the *electric charge*. In the same way, the world of biology is different from the world of chemistry because its description requires at least one new observable, i.e., *information*.

Code Biology, furthermore, has underlined that in living systems we find not only *information* but also *meaning* because any code is a mapping that gives meaning to something. Biological information and biological meaning, in other words, are *observables* because we cannot describe living system without them, and they are *fundamental* observables because they cannot be reduced to other entities.

What is remarkable in this paper is that the author reaches the conclusions of Code Biology coming from Biosemiotics, and this shows how important it is to study the semiotic processes of life with an open mind.

10 – Marcella Faria

Two Theories of Action - Edelman's Neuronal Group Selection and the Poetics of Paul Valéry.

Today the dominant theory in biology is the view that natural selection is the sole mechanism of evolution, a view that is often referred to as *Universal Darwinism*. In the case of the brain, Gerald Edelman has given to this view the name of *Neural Darwinism*, the idea that natural selection is the mechanism that accounts for the basic processes of the nervous system. For a long time, natural selection has been exclusively studied in biological systems, but the author shows that the French poet Paul Valéry reached the conclusion that the works of art too are based on variation and selection. This is the main thesis of the present paper: natural selection is at work not only in the sciences but also in the humanities, and is therefore a bridge between art and science.

Many biologists have pointed out that natural selection is the *major* mechanism of evolution but not the *only* one, and countless proposals have advocated that *something else* is required to account for life. Code Biology is one of them but has two unique features: it has identified what the *something else* actually is, and has proposed that evolution was caused by two different but *equally important* mechanisms.

The discovery of the genetic code has proved that there are *two* distinct molecular mechanisms at the basis of life, the *copying* of the genes and the *coding* of proteins, and these two mechanisms give origin to two distinct mechanisms of evolution: *evolution by natural selection*, based on copying, and *evolution by natural conventions*, based on coding. This is the central concept of Code Biology: coding cannot be reduced to copying, which means that evolution cannot be reduced to natural selection and life cannot be reduced to universal Darwinism.

This paper describes the ideas proposed by Gerald Edelman and Paul Valéry on the key role that natural selection plays in science and in art, and appears therefore to be advocating universal Darwinism, but in reality there is something else in it. The author underlines that Gerald Edelman and Paul Valéry noticed that there are cases where natural selection does not provide a full explanation and this suggests that there may be another mechanism at work in evolution. The author points out that this second mechanism may well be coding and it is possible therefore that the unification of art and science requires not only natural selection but also natural conventions.

11 – João Carlos Major

Archetypes and Code Biology.

It is a widely diffused opinion that the theories proposed by Sigmund Freud and by Carl Gustav Jung have little or no scientific basis so why should we discuss them in Code Biology which is, by its own definition, "*the study of all semiotic processes of life with the standard methods of science*"? The point is that the theories of Freud and Jung were about *dreams*, and even if we dismiss them we still have the problem to understand what dreams are. The extraordinary thing is that we already know the answer and yet most scientists do not seem to be aware of its consequences.

The answer is that dreams are '*manufactured mental processes*', where the key word is '*manufactured*' because this word means '*produced according to rules*', i.e., '*according to codes*'. Dreams, in conclusion, are mental processes that are manufactured according to '*dream codes*'. In reality, the situation is more complicated than that. Dreams are produced by breaking up and by reassembling the same processes that take place in the awake brain, so the '*dream codes*' are in fact '*dissociated brain codes*'. This in turn means that the awake brain is working according to non-dissociated codes, and we have therefore the problem of finding out what these '*brain codes*' are.

Freud and Jung argued that dreams are an Ariadne thread to the mind but this does not exclude the existence of other threads. The author of this paper suggests that mental archetypes are mental codes and this amounts to saying that we can study the unconscious in the framework of Code Biology, a thread that may well be worth thinking about.

4. PART 3 – Codes and semiosis

(Alexander et al., 2021; Gare, 2021; Lacková and Faltýnek, 2021; Zolyan, 2021; Prosdocimi and de Farias, 2021).

12 – Victoria Alexander, Augustus Bacigalupi and Òscar Castro Garcia

Living systems are smarter bots: Slime mold semiosis versus AI symbol manipulation.

The authors examine two recent developments in deep-learning algorithms that appear to go beyond the traditional limitations attributed to Artificial Intelligence. The first is the "*New Rembrandt project*", an algorithm that appears to be capable of producing paintings in the characteristic Rembrandt's style. The second is the computer game "*Hide and Seek*", where the machine can develop winning strategies that the programmers had not anticipated, a result that has been described as an example of *emergent* artificial learning. The authors of this paper, however, show that in both cases the machines can find new ways of achieving a prefixed goal, but they do not spontaneously develop *new* goals and are not capable of *unsupervised learning*, the two faculties which are the hallmarks of animal learning.

The difference between artificial and biological intelligence, furthermore, is used by the authors to reach a highly original conclusion. They notice that living organisms are often described as *machine-like systems*, as if their biological intelligence had the same logic of artificial intelligence. The computer generated game called '*prisoner's dilemma*', for example, is regarded as a real evolutionary strategy, and the whole evolution of species is often described in terms of '*arms race*' as in many computer games.

The conclusion of the paper, in short, is that modern biology is often applying the models of artificial intelligence to describe what happened during the evolution of life on Earth, and this is probably its greatest shortcoming, the victory of artificial biology over real biology.

13 – Arran Gare

Code Biology and the Problem of Emergence.

Arran Gare is the editor-in-chief of the journal '*Cosmos and History*' and in this paper he describes his views on Code Biology as both author and editor. He starts by saying that "*Barbieri has not provided an adequate account of emergence*" and in order to overcome this limitation he calls attention to the contributions of three authors.

The first is Stanley Salthe who "*has clarified to some extent the nature of emergence by conceptualizing it as the interpolation of new enabling constraints. Clearly, codes can be seen as enabling constraints. How this actually happens, though, is still not explained.*"

The second author is Stuart Kauffman who "*has grappled with this issue and shown that it radically challenges the assumptions of mainstream science. He has attempted to reintroduce real possibilities or potentialities into his ontology, and argued that radically new developments in nature are associated with realizing adjacent possibles. This is still not adequate.*"

The third author is the French philosopher Gilbert Simondon, and in this case Arran Gare argues that Simondon has proposed concepts that are directly involved in emergence, in particular the concept of "*transduction as part of ontogenesis in a process of 'individuation', that is, the emergence of individuals from pre-individual fields or milieu.*"

Arran Gare points out that Simondon borrowed the term 'transduction' from Jean Piaget, who distinguished it from the two classical categories of 'induction' and 'deduction' introduced by Aristotle. In reality it was not Piaget who introduced that third category, it was Charles Peirce, and the name that Peirce gave to it was not 'transduction' but 'abduction'. As for the term 'individuation', biology does have a name for the emergence of individuality from pre-individuality, and that name is 'specificity', the 'unique individuality' of the components of all living systems that came into being with the origin of the genetic code. Gilbert Simondon, in other words, gave new names to concepts already known with other names and it is not these tricks that allow us to solve the problem of emergence.

14 – Ľudmila Lacková and Dan Faltýnek

The Lower Threshold as Unifying Principle Between Code Biology and Biosemiotics.

The authors of this paper underline that Code Biology broke away from Biosemiotics on the Peircean concept of interpretation. More precisely, on the idea that the cell is capable of interpreting the world because this would amount to saying that there is a sort of 'mind' in it. The authors argue instead that "*interpretation is not necessarily mental, therefore it is not exclusively related to cognition or mind*". This is a point that keeps coming up in the literature, so let us address it.

Charles Peirce gave a gigantic contributions to philosophy because he introduced in logic the new category of *abduction* in addition to the classical categories of *induction* and *deduction* that Aristotle introduced some 2400 years ago. The philosophical concept of abduction, furthermore, turned out to have a biological counterpart because it corresponds to the mechanism that many animals use to *interpret* what goes on in the world. It is an experimental fact, in other words, that interpretation and abduction are biological processes that take place in the brain. But can we extend them from the brain to the cell?

Protein synthesis is based on the genetic code and any code is a semiotic process, but according to Peirce there cannot be semiosis without interpretation so there must be interpretation in protein synthesis. This amounts to saying that the ribosome scans a messenger RNA and '*interprets*' the information carried by that tape, but this is certainly what is *not* happening. The codons of a messenger RNA are recognized by the complementary anticodons of the transfer-RNAs and saying that a ribosome is interpreting a messenger RNA is like saying that a key is interpreting a lock in order to open a door.

The experimental evidence, in other words, tells us that there is no interpretation in protein synthesis, but the governing board of Biosemiotics insisted that the genetic code *must be* a process of interpretation. The break between Code Biology and Biosemiotics, in conclusion, was not caused by the cognitive abilities of the cell, but by a much stronger argument. It was caused by the fact that Biosemiotics was imposing an interpretation-based view of the genetic code that is in complete contrast with the evidence, and the break was absolutely necessary in order to ensure that the genetic code and all the other semiotic processes of life can be studied as natural phenomena with the standard methods of science.

The authors, on the other hand, point out that despite their differences Biosemiotics and Code Biology have in common the idea that life came into being when the first code appeared on the primitive Earth. This amounts to saying that Biosemiotics and Code Biology agree that the boundary between life and non-life, what the authors call "*the lower threshold*", is the no-man's land of semiosis that some chemical systems had to cross in order to become living systems.

15 – Suren Zolyan

On the context-sensitive grammar of the genetic code.

The discovery of the genetic code has been widely announced as the discovery of a *molecular language*, but most biologists have assumed that this expression is only a linguistic metaphor not a scientific fact. This is because the long dominant *Stereochemical theory* proclaimed that the rules of the genetic code were determined by chemistry and cannot

therefore be arbitrary, whereas a real code is necessarily made of arbitrary rules. In reality, the laboratory experiments have proved beyond doubt that any codon can be associated to any amino acid, which means that the genetic code is really made of arbitrary rules.

There is however another argument against the idea that the genetic code is a molecular language. This is because there does not seem to be a grammar in it. The dictionary of the genetic code contains 64 words, or *codons*, and each word is made of three nucleotides. The nucleotides, in turn, belong to four families and all their 64 possible combinations are functioning codons, which means that all possible combinations of the genetic letters are genetic words. A grammar is a set of rules that allow the formation of *a limited number of words* out of *countless possible combinations of letters*, but in the genetic code all combinations of genetic letters are genetic words and it seems therefore that no grammar exists in it.

The major proposal of this paper is the idea that a genetic grammar does exist but it is a *hidden* one, and consists in the rules by which the nucleotides are arranged in triplets. This is what allows the author to conclude that the genetic code can indeed be regarded as a molecular language.

Another interesting problem comes from this question: let us admit that the genetic code is comparable to a language, but is it comparable to a *written* language or to a *spoken* language? The units of a written language are vowels and consonants whereas the units of a spoken language are the *sounds* (the *phonemes*) of those vowels and consonants, so the question becomes: are the nucleotides equivalent to the *letters* of a written text or to the *sounds* of those letters? The author argues that the genetic code is equivalent to a spoken language not to a written one, and his explanation is another reason that makes the paper worth reading.

16 – Francisco Prosdocimi and Sávio Torres de Farias

Life and living beings under the perspective of organic macrocodes.

Code Biology is based on two key principles: the idea that there are many organic codes in living systems, and the idea that coding is a fundamental mechanism of evolution. The authors of this paper accept these two ideas but they also accept many other ideas and propose a theoretical mixture that in principle goes far beyond Code Biology. In practice, however, the mixture that they propose is not without problems. Three points, in particular, convey this impression.

[1] The authors propose that the primary forms of life were "*Viruses, Archaea and Bacteria*", in line with the view that viruses played a crucial role in the origin of life. The authors suggest that the viral theory of life is fully compatible with Code Biology, but this is in no way a foregone conclusion. What we know is that viruses exploit the codes that already exist in living systems but there is no evidence that they contributed to the evolution of these codes.

[2] The authors accept the idea proposed by Carl Woese (1998) that the first biological systems were *open* in the sense that they had no fixed limitations in space, and propose that these systems (that Woese called *progenotes*) evolved into modern organisms when they became encapsulated by membranes. The authors, however, do not say that the transition from *open* to *close* systems was made possible by new codes – in particular by the *signal transduction codes* – as if codes had nothing to do with it.

[3] The authors point out that life evolved by the emergence of many overlapping codes that are collectively referred to as "*macrocodes*", and this is presented as a principle that applies to all living systems. In reality, the emergence of new codes ended very early in bacteria, and only eukaryotes went on evolving new codes throughout the history of life.

The authors, in conclusion, propose a synthesis of Code Biology with various other theories and in principle this is a reasonable idea. The problem is that there are many different ways of realizing such a synthesis and not all of them provide a convincing explanation of the experimental facts.

5. PART 4 – The genetic Code

(Di Giulio, 2021; Thompson et al., 2021; Pawlak et al., 2021; Fimmel et al., 2021; Konjevoda and Štambuk, 2021).

17 – Massimo Di Giulio

The evolutionary stages of the complexity of biological catalysts mark and clarify the phases of the origin of the genetic code: a model for the origin of the reading frame with codons from proto-mRNAs with different frames.

Di Giulio has a long established reputation in the field of the genetic code and has proposed a variety of unorthodox ideas. In this paper he underlines that the ultimate purpose of the genetic code is the synthesis of proteins and most proteins are enzymes, so he finds it legitimate to assume that enzyme catalysis was the *raison d'être* of the genetic code, the main selective pressure that promoted the evolution of that code. Here, however, some caution is necessary because the evolution of the genetic code was *not* driven by the *future advantages* of the protein enzymes, and it remains to be seen what were the *immediate advantages* that fuelled that evolution.

The model proposed in this paper belongs to a larger view – called the '*coevolution theory*' – that was first proposed by Jeffrey Wong (1975) and was later extended by Di Giulio (2008). This theory states that the evolution of the genetic code went on in parallel with the evolution of metabolism. The key idea, in both the original and the extended form of the coevolution theory, is the hypothesis that on the primitive Earth the synthesis of amino acids was taking place on RNAs, and there was therefore a *metabolic continuity* between RNAs and amino acids, a continuity that would immediately explain the associations between codons and amino acids, i.e., the very rules of the genetic code.

The problem with this theory is that in all known living systems the synthesis of amino acids does *not* take place on RNAs, and there is no convincing evidence that it *could* have taken place on RNAs at the beginning of life. The present paper, in conclusion, is proposing a model that is based on thin ice evidence but it is still worth reading because it gives us an idea of the dramatic problems that life had to face on the primitive Earth at the beginning of its history.

18 – Julie D. Thompson, Raymond Ripp, Claudine Mayer, Olivier Poch and Christian J. Michel

Potential role of the X circular code in the regulation of gene expression.

Eukaryotic protein synthesis takes place in four distinct phases (transcription, splicing, translation and post-translation) and there is at least one code in every one of them: in transcription there is the RNA-polymerase code (described by Giorgio Dieci in this issue); in splicing there are the splicing codes; in translation there is the genetic code and in post-translation there is the histone code. The authors of this paper propose that in translation there is not one but *two* distinct codes: the genetic code and what they call "*the X circular code*". There is a long story behind this idea, so let us make a brief summary.

In protein synthesis the sequence of codons must be read in the correct order to produce the correct sequence of amino acids. A shift of a single nucleotide would produce a completely different sequence of codons and therefore a completely different protein. But how does the cell manage to avoid the frame-shift errors? A solution of this problem was proposed in 1957 by Francis Crick with the idea that the reading of a messenger-RNA can only start with a codon that belongs to a special family of *comma-free codons*, because in this way no shift can occur in the reading frame. In 1961, however, Nirenberg and Matthaei discovered that the codon for phenylalanine is TTT and this is *not* a comma-free

codon, so the comma-free solution had to be abandoned.

In 1996, Arquès and Michel introduced a new class of codes, called "*circular codes*", which are error-correcting and self-synchronizing, properties that allow them to recognize the correct reading frame. The X *circular code* is one of them and consists in 20 codons that code for 12 amino acids, and its mathematical features allow the circular code to retrieve the correct reading frame (Arquès and Michel, 1996).

The X circular code, in other words, avoids the reading-frame errors, but the same was true for the comma-free code and there are, in principle, many other ways to avoid those errors. The problem is that we do not have the experimental evidence that tells us how the cell does cope with the frame-shift errors. The circular code solution, in short, is undoubtedly fascinating but for the time being it is only a theoretical possibility, not an experimental certainty.

Finally, there is one more point that should be taken into account. The rules of the biological codes are always implemented by adaptors, but so far no adaptors have been described in the circular codes and this means that the word '*code*' may not be an appropriate term for them.

19 – Konrad Pawlak, Paweł Błazej, Małgorzata Wnetrzak, Dorota Mackiewicz and Paweł Mackiewicz

Models of genetic code structure evolution with variable number of coded labels.

The genetic code is a set of rules that establish a correspondence between 64 codons and 20 amino acids. More precisely, 61 codons code for amino acids and the remaining 3 code for termination signals. Between 61 codons and 20 amino acids there is a many-to-one correspondence, and this means that the genetic code is *degenerate* (or *redundant*). More precisely, some amino acids are coded by 6 codons, some by 4, others by 2, and only 2 amino acids are coded by a single codon.

Another important fact is that codons are associated to amino acids by transfer-RNAs, and in principle therefore a cell should contain 61 transfer-RNAs for recognizing its 61 codons, but in reality the number of transfer-RNAs in any cell is about 40. The best explanation for this surprising fact is still the proposal made by Francis Crick in 1966 with the *wobble hypothesis*.

Crick pointed out that the three nucleotides of an anticodon stick out like fingers from the surface of the transfer-RNAs and this allows them to oscillate, or *wobble*. The result is that a nucleotide in an anticodon (especially in third position) can form a temporary bond not only with its complementary nucleotide but also with another one with which it has a partial similarity. A transfer-RNA, in other words, can associate the same amino acid to two or more codons, and this explains why the number of transfer-RNAs in a cell is less than the number of its codons.

The authors of the present paper propose a theory that appears to account, in principle, for the above two experimental facts: the degeneracy of the genetic code and the numerical difference that exists between transfer-RNAs and codons. Another interesting point comes from the widely accepted conclusion that during the evolution of the genetic code there has been a systematic increase in the number of amino acids. The problem is that there are 40 transfer-RNAs in any cell, so how do we explain the fact that the number of amino acids went up only to 20 when it could have gone all the way up to 40?

According to the original theory of Code Biology, the evolution of the genetic code was caused by a mechanism that steadily reduced the ambiguity of the ancestral code but such a mechanism requires an increased number of amino acids, and this means that increase came abruptly to an end when the ambiguity of the code was completely eliminated (Barbieri 2019a).

20 – Elena Fimmel, Markus Gumbel, Martin Starman and Lutz Strüngmann

Robustness against point mutations of genetic code extensions under consideration of wobble-like effects.

The computer simulations of the genetic code allow us to generate an unlimited population of *virtual codes* and the study of these codes has given us important clues about the evolution of the real genetic code. It has been found, for example, that the real genetic code performs better than most of its virtual alternatives and this suggests that there have been processes of optimization in the early phases of its evolution. It has been underlined, on the other hand, that there are 10^{84} virtual codes and many of them are more robust than the real code. The computer simulations, in other words, have revealed that the genetic code did go through processes of optimization but apparently it went only half way up the optimization ladder.

The authors of the present paper have used the computer simulation studies in order to find out how robust is the genetic code in respect to point mutations. More precisely they resorted to graph theory and created expanded versions of the genetic code with two operations: (1) the use of tetranucleotides in addition to the standard triplet codons, and (2) the addition of non-canonical base pairs to the genetic alphabet. This allowed them to show that the performance of the genetic code becomes much more robust when weights are introduced in the point mutations that mimic the wobble effect.

The major result of the computer simulation studies, as we have seen, is the conclusion that the genetic code went only half way up the optimization ladder, and the original theory of Code Biology does provide an explanation for this result. According to that theory, the evolution of the genetic code was caused by a mechanism that steadily reduced the ambiguity of the ancestral code, but any step that reduced ambiguity was also a step that contributed to the optimization process, which means that this process came abruptly to an end when the ambiguity of the code was completely eliminated (Barbieri 2019a).

Perhaps one day the computer simulation studies will give us a model of the ambiguity reduction mechanism and make us understand the logic that was behind the evolution the genetic code.

21 – Paško Konjevoda and Nikola Štambuk

Relational Model of the Standard Genetic Code.

The classical table of the genetic code has been the same since the 1960s and so far nobody has proposed to change it but this is precisely what the authors of the present paper are doing. They are proposing to replace that classical table – that represents the meanings of 64 codons – with 4 smaller tables that represent 16 codons each, a description that they call the "*Relational Model*" of the genetic code. The question that comes immediately to mind is "*why?*" and to this the authors give not one but two distinct answers.

The first reason for adopting the Relational Model is the fact that model is the best known way for storing and retrieving data. The model was first proposed in 1970 by the British computer scientist Edgar Codd and even today, after more than 50 years, it is the most widely used method of data storage and management. The model is built in stages, called *normalization* steps, where each step eliminates useless data and optimizes the performance. From this point of view, the classical table of the genetic code is a *non-normalized* description, whereas the four tables of the relational model are *fully normalized* and are therefore a more reliable representation of the code.

The second reason for adopting the Relational Model comes from the fact that this model appears to have a straightforward biological implication: it seems to suggest that the modern genetic code with 64 triplet codons evolved from the fusion of several simpler codes based on 16 doublet codons.

This is a fascinating conclusion, but in reality it is compatible with two different scenarios. One is that the Relational Model represents what historically happened in the evolution of the genetic code, and is therefore a sort of *frozen history* of that code. The other is that the history of the genetic code was somehow *constrained* by some underlying relationships and the Relational Model tells us what those relationships

were.

This is the great merit of the paper: it gives us an entirely new framework for studying the history and the underlying principles of the genetic code.

References

Alexander, V.N., Bacigalupi, J.A., Castro Garcia, Ò.C., 2021. Living systems are smarter bots: Slime mold semiosis versus AI symbol manipulation. Biosystems 206, 104430. https://doi.org/10.1016/j.biosystems.2021.104430.

Arquès, D.G., Michel, C.J., 1996. A complementary circular code in the protein-coding genes. J. Theor. Biol. 182, 45–58. https://doi.org/10.1006/jtbi.1996.0142.

Barbieri, M., 1981. The ribotype theory on the origin of life. J. Theor. Biol. 91, 545–601. https://doi.org/10.1016/0022-5193(81)90211-3.

Barbieri, M., 1985. The Semantic Theory of Evolution. Harwood Academic Publishers, London & New York. Reissued in 2020 in the *Routledge Library Editions*. https://www.taylorfrancis.com/books/9780429290039.

Barbieri, M., 2003. The Organic Codes. An Introduction to Semantic Biology. Cambridge University Press, Cambridge, UK.

Barbieri, M., 2012. Codepoiesis – the deep logic of life. Biosemiotics 5, 297–299. https://doi.org/10.1007/s12304-012-9162-4.

Barbieri, M., 2014. From biosemiotics to code biology. Biol. Theory 9, 239–249. https://doi.org/10.1007/s13752-013-0155-6.

Barbieri, M., 2015. Code Biology. A New Science of Life. Springer, Dordrecht.

Barbieri, M., 2016. From the common ancestor to the first cells: the code theory. Biol. Theory 11, 102–112. https://doi.org/10.1007/s13752-016-0236-4.

Barbieri, M., 2017. How did the eukaryotes evolve? Biol. Theory 12, 13–26. https://doi.org/10.1007/s13752-016-0253-3.

Barbieri, M., 2018. What is code biology? Biosystems 164, 1–10. https://doi.org/10.1016/j.biosystems.2017.10.00.

Barbieri, M., 2019a. Evolution of the genetic code. The ambiguity-reduction theory. Biosystems 185, 104024. https://doi.org/10.1016/j.biosystems.2019.104024.

Barbieri, M., 2019b. Code biology, Peircean biosemiotics and Rosen's relational biology. Biol. Theory 14, 21–29. https://doi.org/10.1007/s13752-018-0312-z.

Barbieri, M., 2020. The semantic theory of language. Biosystems 190, 104100. https://doi.org/10.1016/j.biosystems.2020.104100.

Coca, J.R., Eraña, H., Castilla, J., 2021. Biosemiotics comprehension of PrP code and prion disease. Biosystems 210, 104542. https://doi.org/10.1016/j.biosystems.2021.104542.

Di Giulio, M., 2008. An extension of the coevolution theory of the origin of the genetic code. Biol. Direct 3, 37. https://doi.org/10.1186/1745-6150-3-37.

Di Giulio, M., 2021. The evolutionary stages of the complexity of biological catalysts mark and clarify the phases of the origin of the genetic code: a model for the origin of the reading frame with codons from proto-mRNAs with different frames. Biosystems 207, 104449. https://doi.org/10.1016/j.biosystems.2021.104449.

Dieci, G., 2021. Removing quote marks from the RNA polymerase II CTD 'code'. Biosystems 207, 104468. https://doi.org/10.1016/j.biosystems.2021.104468.

Faria, M., 2021. Two theories of action – Edelman's neuronal group selection and the poetics of Paul Valéry. Biosystems, 104549. https://doi.org/10.1016/j.biosystems.2021.104549.

Fimmel, E., Gumbel, M., Starman, M., Strüngmann, L., 2021. Robustness against point mutations of genetic code extensions under consideration of wobble-like effects. Biosystems 208, 104485. https://doi.org/10.1016/j.biosystems.2021.104485.

Fröhlich, H., 1970. Long range coherence and the action of enzymes. Nature 228, 1093. https://doi.org/10.1038/2281093a0.

Gare, A., 2021. Code biology and the problem of emergence. Biosystems 208, 104487. https://doi.org/10.1016/j.biosystems.2021.104487.

Heng, J., Heng, H.H., 2021. Karyotype coding: the creation and maintenance of system information for complexity and biodiversity. Biosystems 208, 104476. https://doi.org/10.1016/j.biosystems.2021.104476.

Hofmeyr, J.-H.C., 2021. A biochemically-realisable model of the self-manufacturing cell. Biosystems 207, 104463. https://doi.org/10.1016/j.biosystems.2021.104463.

Igamberdiev, A.U., 2021. The drawbridge of nature: evolutionary complexification as a generation and novel interpretation of coding systems. Biosystems 207, 104454. https://doi.org/10.1016/j.biosystems.2021.104454.

Konjevoda, P., Štambuk, N., 2021. Relational model of the standard genetic code. Biosystems 210, 104529. https://doi.org/10.1016/j.biosystems.2021.104529.

Kun, Á., 2021. The major evolutionary transitions and codes of life. Biosystems, 104548. https://doi.org/10.1016/j.biosystems.2021.104548.

Kühn, S., Hofmeyr, J.-H.C., 2014. Is the "histone code" an organic code? Biosemiotics 7, 203–222. https://doi.org/10.1007/s12304-014-9211-2.

Lacková, L., Faltýnek, D., 2021. The lower threshold as unifying principle between code biology and biosemiotics. Biosystems 210, 104523. https://doi.org/10.1016/j.biosystems.2021.104523.

Major, J.C., 2021. Archetypes and code biology. Biosystems 208, 104501. https://doi.org/10.1016/j.biosystems.2021.104501.

Maturana, H.R., Varela, F., 1980. Autopoiesis and Cognition: the Realisation of the Living. D. Reidel Publishing Company, Dordrecht.

Maynard Smith, J., Szathmáry, E., 1995. The Major Transitions in Evolution. Oxford University Press, Oxford.

Paredes, O., Morales, A., Mendizabal, A.P., Romo-Vázqueza, R., 2021. Metacode: one code to rule them all. Biosystems 208, 104486. https://doi.org/10.1016/j.biosystems.2021.104486.

Pawlak, K., Błazej, P., Wnetrzak, M., Mackiewicz, D., Mackiewicz, P., 2021. Models of genetic code structure evolution with variable number of coded labels. Biosystems 210, 104528. https://doi.org/10.1016/j.biosystems.2021.104528.

Petoukhov, S.V., 2021. Algebraic harmony and probabilities in genomes. Long-range coherence in quantum code biology. Biosystems 209, 104503. https://doi.org/10.1016/j.biosystems.2021.104503.

Prosdocimi, F., de Farias, S.T., 2021. Life and living beings under the perspective of organic macrocodes. Biosystems 206, 104445. https://doi.org/10.1016/j.biosystems.2021.104445.

Prusiner, S.B., 1982. Novel proteinaceous infectious particles cause scrapie. Science 216 (4542), 136–144. https://doi.org/10.1126/science.6801762.

Thompson, J.D., Ripp, R., Mayer, C., Poch, O., Michel, C.J., 2021. Potential role of the X circular code in the regulation of gene expression. Biosystems 203, 104368. https://doi.org/10.1016/j.biosystems.2021.104368.

Woese, C., 1998. The universal ancestor. Proc. Natl. Acad. Sci. U.S.A. 95 (12), 6854–6859. https://doi.org/10.1073/pnas.95.12.6854.

Wong, J.T., 1975. A co-evolution theory of the genetic code. Proc. Natl. Acad. Sci. U.S.A. 72 https://doi.org/10.1073/pnas.72.5.1909, 1909–1912.

Zámečník, L., 2021. Causal and non-causal explanations in code biology. Biosystems 209, 104499. https://doi.org/10.1016/j.biosystems.2021.104499.

Zolyan, S., 2021. On the context-sensitive grammar of the genetic code. Biosystems 208, 104497. https://doi.org/10.1016/j.biosystems.2021.104497.

Marcello Barbieri
Dipartimento di Morfologia Ed Embriologia, Via Fossato di Mortara 64a, 44121, Ferrara, Italy
E-mail address: brr@unife.it.