BioSystems 164 (2018) 1–10

Contents lists available at ScienceDirect

BioSystems

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

Review article

What is code biology?

Marcello Barbieri

Dipartimento di Morfologia ed Embriologia, Via Fossato di Mortara 64a, 44121 Ferrara, Italy

ARTICLE INFO

Article history:
Received 27 July 2017
Received in revised form 4 October 2017
Accepted 5 October 2017
Available online 6 October 2017

Keywords:
Organic codes
Complexity
Macroevolution
Common ancestor
Cell organization
Prokaryotes
Eukaryotes

ABSTRACT

Various independent discoveries have shown that many organic codes exist in living systems, and this implies that they came into being during the history of life and contributed to that history. The genetic code appeared in a population of primitive systems that has been referred to as the *common ancestor*, and it has been proposed that three distinct signal processing codes gave origin to the three primary kingdoms of Archaea, Bacteria and Eukarya. After the genetic code and the signal processing codes, on the other hand, only the ancestors of the eukaryotes continued to explore the coding space and gave origin to splicing codes, histone code, tubulin code, compartment codes and many others. A first theoretical consequence of this historical fact is the idea that the Eukarya became increasingly more complex because they maintained the potential to bring new organic codes into existence. A second theoretical consequence comes from the fact that the evolution of the individual rules of a code can take an extremely long time, but the origin of a new organic code corresponds to the appearance of *a complete set of rules* and from a geological point of view this amounts to *a sudden* event. The great discontinuities of the history of life, in other words, can be explained as the result of the appearance of new codes. A third theoretical consequence comes from the fact that the organic codes have been highly conserved in evolution, which shows that they are the great invariants of life, the sole entities that have gone intact through billions of years while everything else has changed. This tells us that the organic codes are fundamental components of life and their study – the new research field of *Code Biology* – is destined to become an increasingly relevant part of the life sciences.

Contents

E-mail address: brr@unife.it

https://doi.org/10.1016/j.biosystems.2017.10.005
0303-2647/

1. The ontological problem

The *International Society of Code Biology* was established in 2012 and in its constitution it is written that "*Code Biology is the study of all codes of life, from the genetic code to the codes of culture, with the standard methods of science*" (Barbieri, 2014). Code Biology is therefore a well-defined field of scientific research and like many other fields is made of two complementary parts, one experimental and one theoretical.

The experimental part is due to the fact that various independent discoveries in the recent past have shown that many organic codes exist in living systems *in addition* to the genetic code (Barbieri, 2003, 2015a). The theoretical part comes from the fact that these discoveries give new answers to some of the great problems of biology, in particular to the problems of biological complexity, of cell organization and of the causes of macroevolution.

Before discussing these problems, however, we must address a preliminary one. We must face the *ontological problem* of the reality of the organic codes: are they real codes? Do they actually exist in living systems? It is a fact that the genetic code has been universally accepted into Modern Biology, but let us not be naive about this: what has been accepted is the *name* of the genetic code, not its *ontological reality*.

More precisely, the genetic code has been accepted under the assumption that its rules were determined by chemistry and do not have the *arbitrariness* that is essential in any real code. The theoretical premise of this assumption is the belief that there cannot be arbitrary rules in Nature, and this inevitably implies that the genetic code is a metaphorical entity, not a real code. This idea has a long history and let us not forget that for many decades it has been the dominant view in molecular biology.

The very first model of the genetic code was the *Stereochemical Theory*, an idea proposed by George Gamow in 1954 (Gamow, 1954) and later re-proposed by many other authors, which states that the relationships between codons and amino acids, were the result of stereochemical affinities. The second canonical model was the *Coevolution Theory* proposed by Wong (1975, 1981), according to which the genetic code coevolved with the biochemical pathways that introduced new amino acids in protein synthesis. Both theories imply that the rules of the genetic code are chemically determined and no arbitrariness exists in them.

It has taken a long time and much experimental work to overturn this conclusion, but eventually it has been shown that there is no deterministic link between codons and amino acids because any codon can be associated with any amino acid (Schimmel, 1987; Hou and Schimmel, 1988; Budisa, 2004; Hartman et al., 2007; Ling et al., 2015; Acevedo-Rocha and Budisa, 2016). This means that the rules of the genetic code do not descend from chemical necessity and in this sense they are arbitrary.

Today, in other words, we have the experimental evidence that the genetic code is a real code, a code that is compatible with the laws of physics and chemistry but is not dictated by them. Our problem, therefore, is to take stock of this reality and to account for it. How do we explain the existence of arbitrary rules in Nature? How could arbitrary rules evolve on the primitive Earth?

2. The first genetic code

The ribosomal RNAs are among the most conserved molecules in evolution and have the ability to form peptide bonds (Nitta et al., 1998), which means that they appeared very early on the primitive Earth and could stick amino acids together into peptides and polypeptides.

The transfer-RNAs too have been highly conserved in evolution, and their function is to carry amino acids to the ribosomal RNAs, which means that they too became involved very early in protein synthesis. A primitive apparatus, on the other hand, required also a third type of RNAs. When the amino acids are delivered to the site of synthesis, it is necessary that they are kept at a close distance for a long enough time to allow the formation of a peptide bond (Wolf and Koonin, 2007; Fox, 2010). This means that the transfer-RNAs must have temporary anchoring sites, and in primitive systems these were provided by *anchoring RNAs*, the ancestors of the modern *messenger-RNAs* (Osawa, 1995).

We do not know *how* the three types of RNAs came into being, but it is a *historical fact* that they did because they exist in every cell and this implies that they were present in the common ancestor of all living creatures. It is a historical fact, in other words, that the common ancestor had a primitive apparatus of protein synthesis where the transfer-RNAs were delivering amino acids to the ribosomal-RNAs in the order provided by the codons of the anchoring RNAs. This apparatus was automatically creating a mapping between codons and amino acids, and any such mapping is, by definition, a *genetic code*.

This is how the first genetic code appeared on Earth. It came into being when primitive ribosomal-RNAs, primitive transfer-RNAs and primitive anchoring-RNAs came together in the common ancestor and formed the first apparatus of protein synthesis. But what kind of proteins were manufactured by that apparatus? Here is a problem where the answer is strongly dependent on theory.

According to the stereochemical theory and to the coevolution theory, the evolution of the genetic code consisted primarily in the introduction of new amino acids in protein synthesis, under the assumption that by chemical necessity every codon was translated into one and only one amino acid. This amount to saying that no ambiguity ever existed in the genetic code and all proteins manufactured in the common ancestor were *specific* proteins.

The opposite is true if one accepts that the rules of the genetic code were arbitrary. In this case, there were no necessary relationships between codons and amino acids and the apparatus of protein synthesis could not produce specific proteins. In 1965, Carl Woese made a vivid reconstruction of the situation in the common ancestor: " *the translation mechanism was a far more rudimentary thing than at present, in particular far more prone to make translation errors... We shall assume that errors in translation were extreme, to such an extent that the probability of translating correctly any given messenger-RNA was essentially zero. From this concept of error-ridden translation in the primitive cell it follows that the proteins produced by any given gene will have to be statistical proteins.*" (Woese, 1965, pp. 1548–1549).

From this premise Woese proposed an "*error-minimization theory*" according to which "*... the evolution of the genetic code starts with a primitive cell possessing random, ambiguous codon assignments, and a very error-ridden translation process, and it was the 'necessity' of minimizing the effects of translation errors that led to the highly ordered code that we observe today.*" (Woese 1965, p.1550).

Different versions of the error-minimization theory have been proposed after that, and the overall result, summarized in 2009 by Koonin and Novozhilov, is that there are three major groups of theories on the evolution of the genetic code: the *stereochemical theory*, the *coevolution theory* and the *error-minimization theory* (Koonin and Novozhilov, 2009). In reality, however, a fourth possibility does exist.

3. Evolution of an ambiguous code

A genetic code is *ambiguous* if a codon is translated into two or more amino acids, and it has been pointed out that the first genetic codes that appeared on Earth were necessarily ambiguous (Fitch

and Upper, 1987; Osawa, 1995). This means that a sequence of codons was translated some time into a protein and some other time into a different protein, and the ancestral apparatus of protein synthesis was inevitably producing *statistical proteins*.

It must be underlined that the production of statistical proteins *by code ambiguity* is substantially different from the production of statistical proteins *by translation errors*. In the first case, the system is intrinsically unable to produce specific proteins, whereas in the second case the system can produce specific proteins but these are turned into statistical proteins by the translation errors.

The common ancestor, in other words, contained two different types of statistical proteins: those produced by code ambiguity and those produced by translation errors. This implies that the evolution of the genetic code took place by two distinct mechanisms, one that reduced code ambiguity and one that minimized the effects of translation errors. For a long time, the evolution of the genetic code by ambiguity-reduction has only been an abstract possibility, but recently one mechanism (*the ribosome-oriented model*) has been proposed (Barbieri, 2015b).

The idea that the first genetic code was ambiguous implies that all proteins of the common ancestor were statistical proteins, and this immediately raises a problem: how could the primitive ribosomes work with statistical proteins? It has been shown that the key functions in protein synthesis are performed by the RNAs, not by proteins, and this suggests that ribosomes do not necessarily need specific ribosomal proteins.

There are also other clues that point in that direction: (1) the ribosomal proteins in bacteria have a smaller molecular weight than in eukaryotes, and this means that their molecular weight is not a critical factor; (2) the ribosomal proteins in bacteria are fewer that those in eukaryotes, and this means that their number is not a critical factor; (3) the ribosomal proteins can differ from species to species, and this means that their composition is not a critical factor.

The ribosomal proteins, in short, can have different molecular weights, different numbers and different compositions, and yet they all give origin to fully operational ribosomes. This means that their functions do not require specific proteins, but only molecules that collectively belong to a family. In particular, it means that the ribosomal proteins of the common ancestor could well have been families of statistical proteins. There was however one condition that had to be met.

The ribosomes are formed by self-assembly from their components, and their evolution requires that they are re-assembled in every generation with characteristics that are similar to those of the previous generation. In the ancestral systems this was possible only if the statistical differences that existed in each family of ribosomal proteins were not cancelled out by the ambiguity of the genetic code and could reappear in every new generation.

The ambiguity of the code, in other words, was the limiting factor that determined how many families of statistical ribosomal proteins could reappear in the descendants. Which means that only by lowering the ambiguity of the genetic code it was possible to increase the number of ribosomal proteins, to diversify their functions and to improve their performance in ribosome self-assembly.

It will be noticed that the evolution of the ribosomal proteins was not about this or that protein or this or that protein function. It was involving all ribosomal proteins and all ribosomal functions at the same time. The rules of the genetic code were arbitrary in the sense that anyone of them could have been different because what counted were not their *individual* features but their *collective* relationships. It was the evolution of a systems as a whole, an evolution that went on until the ambiguity of the genetic code was completely removed and *biological specificity* in protein synthesis came into existence.

4. The modern genetic code

In the modern apparatus of protein synthesis the 20 canonical amino acids are carried by 20 aminoacyl-tRNA-synthetases, each of which attaches one amino acid to one or more tRNAs. The synthetases are specific proteins that can be produced only by an apparatus that already has a genetic code, and this gives us a classic *chicken-and-egg* paradox: how could the synthetases be brought into existence if their production required an apparatus of protein synthesis in which they were already present?

A possible solution is that the *modern* apparatus of protein synthesis was preceded by an *ancient* apparatus where the amino acids were carried not by protein-synthetases but by RNAs (Maizels and Weiner, 1987). The *modern genetic code*, in other words, was preceded by an *ancient genetic code* based on RNA-synthetases that were later replaced by proteins.

Such a replacement, on the other hand, was bound to have major consequences. Proteins are substantially different from RNAs, and replacing the RNA synthetases with protein synthetases was bound to modify the rules of the ancient genetic code. But do we have any evidence that such change actually took place?

The idea that the rules of the ancient genetic code were repeatedly modified has received a strong support from a variety of computer studies which have shown that the modern genetic code performs better than most of its many potential alternatives (Haig and Hurst, 1991; Freeland and Hurst, 1998; Bollenbach et al., 2007). The very fact that the genetic code went through an optimization phase implies that its first rules did change and this means that the ancient genetic code was effectively replaced by the modern one.

The optimization of the genetic code has often been described as a novel evolutionary mechanism, but in reality it is only a different way of describing an error-minimization mechanism. According to Carl Woese, for example, the evolution of the genetic code was driven by the need to minimize the translation errors (Woese, 1965; Woese et al., 1966), and this is equivalent to saying that it was driven by the need to optimize the results of translation. According to Gilis et al. (2001) the modern code was optimized in respect to the stabilization of protein structure, and this is equivalent to saying that its evolution tended to minimize the deleterious effects of unstable protein structures.

The evidence that we have, in conclusion, suggests that there have been two distinct evolutions of the genetic code: one that reduced the ambiguity of the first genetic codes and one that improved the efficiency of the translation apparatus until the point was reached when the accuracy of protein synthesis became so high as to be virtually error-free.

We have no direct evidence of those two evolutions, but it is worth considering the idea that both of them were *ribosome-oriented*: the idea that the evolution of the ancient genetic code went hand in hand with the evolution of the ribosomal proteins, whereas the evolution of the modern genetic code went hand in hand with the introduction of modern synthetases in the apparatus of protein synthesis (Barbieri, 2015b).

5. From the common ancestor to the first cells

The phylogenetic trees reconstructed from molecular data have shown that all living creatures belong to three primary kingdoms, or domains, that are referred to as Archaea, Bacteria and Eukaya, and the first cells that appeared on Earth were the first representatives of these kingdoms (Woese and Fox, 1977; Woese et al., 1990; Woese, 2000). The three types of cell share a few universal features that they received from the common ancestor, but most of their characteristics are unique to each kingdom, which means that they evolved independently in the descendants of the common ancestor. The cells of the three kingdoms, in particular, have three distinct

types of cell membrane, and this gives us a major evolutionary problem.

The cell membrane is the site where molecules are transported to and from the environment (*molecular transport*), and where energy is obtained from outside sources and converted into internal forms (*energy transduction*). These two processes – the exchange of matter and the access to energy – are so fundamental that we can hardly imagine a common ancestor without them, and yet the phylogenetic data tell us in no uncertain terms that ancestral membrane was not conserved and the descendants of the common ancestor evolved independently three different types of membranes.

In order to deal with this problem, let us underline that the cell membrane is also the site of *signal transduction*, the process that transforms the signals from the environment (*first messengers*) into internal signals (*second messengers*). First and second messengers belong to two independent worlds and laboratory experiments have shown that the same first messenger can activate different second messengers and that different first messengers can activate the same second messenger (Alberts et al., 2007) which means that there are no necessary connections between them.

The membrane receptors that implement signal transduction, furthermore, are molecular adaptors that create links between first and second messengers just as the transfer-RNAs create links between codons and amino acids. In signal transduction, in short, we find all the essential components of a code: (a) two independents worlds of molecules (first messengers and second messengers), (b) a set of adaptors that create a mapping between them, and (c) the proof that the mapping is arbitrary because its rules can be changed in many different ways. All of this amounts to saying that signal transduction is based on *signal transduction codes* (Barbieri, 2003), but what was the role of these codes in evolution?

It is a historical fact that the origin of the genetic code was a major turning point and yet it was not enough to create a modern cell. The reason is that the descendants of the common ancestor could produce specific proteins but not *specific responses to the environment* because they had not yet evolved a modern signal transduction system. They had biological specificity in protein synthesis, but not in their interactions with the world.

This allows us to formulate a hypothesis on the role of the signal transduction codes in evolution. As the genetic code was instrumental to the transition from statistical to specific proteins, the signal transduction codes were instrumental to the transition from statistical to specific cell behaviours, and it is precisely the appearance of specific cell characteristics that we associate with the origin of the first modern cells (Barbieri, 2016). It is an experimental fact, furthermore, that Archaea, Bacteria and Eucarya have three different types of membranes and three distinct signalling systems (Harold, 2014). This suggests that the descendants of the last common ancestor evolved along independent lines and gave origin to three distinct types of cells by combining the universal genetic code with three distinct signal-transduction codes.

6. A multitude of organic codes

In addition to the genetic code it has been found that many other organic codes exist in living systems. Among them, the *metabolic code* (Tomkins, 1975), the *sequence codes* (Trifonov, 1989, 1996, 1999), the *histone code* (Strahl and Allis, 2000; Turner, 2000, 2002, 2007; Kühn and Hofmeyr, 2014), the *sugar code* (Gabius, 2000, 2009), the *splicing codes* (Barbieri, 2003; Fu, 2004; Buratti et al., 2006; Wang and Cooper, 2007; Wang and Burge, 2008; Solis et al., 2008; Cooper et al., 2009; Tazi et al., 2009), the *compartment codes* (Barbieri, 2003), the *cytoskeleton code* (Barbieri, 2003; Gimona, 2008), the *tubulin code* (Verhey and Gaertig, 2007; Janke, 2014; Raunser and Gatsoiannis, 2015; Barisic and Maiato, 2016; Chakraborti et al., 2016), the *nuclear signalling code* (Maraldi, 2008), the *injective organic codes* (De Beule et al., 2011; De Beule, 2014), the *molecular codes* (Görlich et al., 2011; Görlich and Dittrich, 2013), the *ubiquitin code* (Komander and Rape, 2012) and the *glycomic code* (Buckeridge and De Souza, 2014).

It must be underlined that the discovery of the genetic code has been facilitated by two particularly favourable features. More precisely, by the fact that (1) the adaptors are single molecules (the tRNAs) and (2) the components of the code form a closed set (64 codons and 20 amino acids). In many other organic codes, instead, the situation is more complicated because the adaptors can be combinations of molecules (*combinatorial codes*), and the domain (or *alphabet*) of the codes can be open and potentially unlimited. Let us illustrate the concept of combinatorial codes with two examples.

[1] In the case of *splicing*, the existence of a code is suggested by the fact that the splicing machines, known as *spliceosomes*, are huge molecular structures like ribosomes, and that splicing employs small molecules, called *small nuclear RNAs* (*snRNA*), that are comparable to transfer-RNAs. What actually proves the existence of spicing codes is that the snRNAs are real adaptors because they perform two independent recognition processes, one for the beginning and the other for the end of the spliced pieces.

It turned out however that there are two major complications in splicing in respect to protein synthesis. One is the fact that the order in which the exons are joined together can be shuffled in various ways, an operation, called *alternative splicing*, that allows the cell to obtain a whole family of proteins from the same gene. The other complication is the fact that many introns carry sequences that are similar to exons but translate into nonsense and for this reason are called *pseudo exons* or *pseudo genes*. They would create havoc if incorporated into mRNAs and the splicing machinery had to evolve the means to differentiate real exons from pseudo ones. The result is that real exons contain internal identity marks that are known as *exonic splicing enhancers* (ESEs) and *exonic splicing silencers* (ESSs) (Fu, 2004; Matlin et al., 2005; Pertea et al., 2007; Barash et al., 2010; Dhir et al., 2010). The presence of these marks, in turn, means that the adaptors of the splicing codes are not single molecules but combinations of molecules because they must be able to recognize not only the beginning and the end of the real exons, but also their internal identity marks.

[2] Another outstanding example of combinatorial code is the *histone code*, due to the discovery that the histone modifications do not act individually and their final result is due to a *combination* of histone marks rather than a single one. This led David Allis and Bryan Turner to propose that the histone marks operate in combinatorial groups, like letters that are put together into the words of a molecular 'language' that was referred to as *histone code* (Strahl and Allis, 2000; Turner, 2000, 2002, 2007; Berger, 2007).

7. Two opposite types of evolution

The reconstruction of the molecular tree of life was obtained at first by studying the differences that exist between individual molecules in different species (Zuckerkandl and Pauling, 1965; Woese and Fox, 1977), but a much more powerful approach became possible by comparing the differences that exist between entire genomes (Koonin, 2003; Snel et al., 2005; Simonson et al., 2005; Jun et al., 2010).

One of the most important results of this extended technology was the discovery that all modern eukaryotes belong to 5 or 6 major groups that radiated from a common ancestor (Baldauf, 2003; Adl et al., 2005; Keeling et al., 2005; Koonin, 2012). This tells us that there have been two major events in the evolution of the cells. The first was the appearance of a population of primitive systems that

evolved the genetic code and has become known as the *Last Universal Common Ancestor* (LUCA); the other was the appearance of the *Last Eukaryotic Common Ancestor* (LECA) the population from which all modern eukaryotic organisms have descended.

The universal ancestor appeared around 3.5 billion years ago, whereas the eukaryotic ancestor came into being roughly 2 billion years later, about 1.5 billion years ago (Harold, 2014). The crucial point is that throughout that period the evolution of the cells took place in two completely different way.

The fossil record has revealed the presence of fossilized bacteria in Precambrian rocks, and has shown that the stromatolites built by cyanobacteria two and three billion years ago are virtually identical to those built by their modern descendants (Barghoorn and Tyler, 1965; Knoll, 2003). The bacteria, in other words, appeared very early in the history of life and have conserved their complexity (in terms of size, shape and number of components) ever since. This point has been beautifully illustrated by Nick Lane: "*... the bacteria and archaea have barely changed in 4 billion years of evolution. There have been massive environmental upheavals in that time. The rise of oxygen in the air and oceans transformed environmental opportunities, but the bacteria remained unchanged. Glaciations on a global scale (snowball earths) must have pushed ecosystems to the brink of collapse, yet bacteria remained unchanged. . ..Nothing is more conservative than a bacterium*" (Lane, 2015, p. 158).

The eukaryotes, instead, did the very opposite. They repeatedly increased the complexity of their cells and eventually broke the cellular barrier and gave origin to plants and animals, to nervous systems and mind, in short to all living creatures that we see around us. This gives us one of the major problems in evolution: why have the prokaryotes *not* increased their complexity throughout the history of life while the eukaryotes have become increasingly more complex?

An unexpected solution to this problem has come from the discovery that the eukaryotes evolved many more organic codes than prokaryotes. The common ancestor gave origin to the genetic code and transmitted to its descendants not only the ability to conserve this code, but also the potential to evolve the new codes that we find in Archaea, Bacteria and Eukarya.

The key point is that the organic codes are key components of the great apparatuses of the cell (the translation apparatus, the signal transduction system, the splicing apparatus, the cytoskeleton, the histone system and the subcellular compartments), and this *experimental fact* does suggest a solution to the problem of complexity. It suggests that the prokaryotes did not increase their complexity because they did not evolve new organic codes whereas the eukaryotes became increasingly more complex because they maintained the potential to bring new organic codes into being. But how did that happen?

8. Complexity and streamlining

In prokaryotes there are only a few organic codes whereas many more are found in eukaryotes, and this does have implications. It suggests that at a very early stage of evolution the prokaryotes lost the *potential* to evolve new organic codes and only the ancestors of the eukaryotes (the cells that Carl Woese called *urkaryotes*) maintained that potential. A natural explanation of this difference is suggested by the fact that the prokaryotes became increasingly committed to fast replication and adopted a drastic *streamlining strategy* in order to achieve that goal. Let us illustrate this point with two examples.

In bacteria, the transcription of the genes is immediately followed by their translation into proteins, but such a fast link is not a primitive feature and could hardly have been present in the common ancestor when the genetic code was still evolving. A direct coupling between transcription and translation required the abolition of all intermediate steps and could be achieved only by the descendants of the common ancestor that adopted a streamlining strategy. The other descendants maintained a physical separation between transcription and translation and this allowed them to gradually introduce in between the operations of splicing. The prokaryotes, in other words, could not evolve a splicing code simply because they had abolished the separation between transcription and translation that is the very precondition of splicing.

As a second example let us consider the histone code. The ancestral DNAs were negatively charged molecules that inevitably attracted positively charged ones, but in order to maximize the replication rate it was necessary to remove any interposition of material between genes and signalling molecules, and this is why the streamlining strategy produced genes with no protein wrapping around them. The ancestral systems that did not follow that strategy, on the other hand, continued to carry genes surrounded by positively charged molecules and eventually some of these evolved into histones. The potential to evolve the histone code, in other words, survived only in the descendants of the common ancestor that did not adopt the streamlining strategy of the bacteria.

We have in this way a solution to the problem of complexity: the cells that adopted a streamlining strategy lost the potential to evolve new organic codes and have conserved the same complexity ever since; the cells that did not adopt a streamlining strategy maintained the potential to evolve new organic codes and gave origin to increasingly complex systems (Barbieri, 2017).

The existence of many organic codes, in conclusion, tells us that they appeared throughout the history of life and that each of them added a new level of complexity to the evolving cells. The complexity of the cell, in other words, is due not only to the number of its components, but also (and perhaps primarily) to the number of rules that exist between its components, i.e., to the number of its organic codes.

9. Codes in multicellular systems

The evolution of life took place exclusively in single cells for about three billion years, but eventually some eukaryotes gave origin to multicellular creatures and again we find that new levels of complexity were associated with new organic codes. Among them: the *Hox* code (Paul Hunt et al., 1991; Kessel and Gruss, 1991), the *adhesive code* (Redies and Takeichi, 1996; Shapiro and Colman, 1999), the *transcriptional codes* (Jessell, 2000; Marquard and Pfaff, 2001; Ruiz i Altaba et al., 2003; Flames et al., 2007), the *apoptosis code* (Basañez and Hardwick, 2008; Füllgrabe et al., 2010), the *bioelectric code* (Tseng and Levin, 2013; Levin, 2014) and the *acoustic codes* (Farina and Pieretti, 2014).

With the origin of the nervous systems, a new type of entities came into being; entities such as *action potentials* and *neuron firings* at the cellular level, and entities such as *instincts and feelings* at the organism level. In addition to organic evolution, based on organic molecules, what came into being was *neural evolution*, based on neural states. The difference between them comes from the fact that organic molecules are *space-objects*, in the sense that their properties come from their three-dimensional organization in space, whereas neural states are *time-objects* in the sense that they arise from sequences of neuron firings in time.

The animal sense organs transform external signals into neural impulses, and are comparable to the signal transduction receptors of the cell that transform first messengers into second messengers. In both cases there is no necessary connection between external and internal signals and the transductions take place according to coding rules. It has been discovered, for example, a *neural code* for the transmission of mechanical stimuli (Nicolelis and Ribeiro, 2006;

Nicolelis, 2011), a *neural code for taste* in the gustatory system (Di Lorenzo, 2000; Hallock and Di Lorenzo, 2006), an *odor receptor code* in the olfactory system (Dudai, 1999; Ray et al., 2006), a *synaptic code* for cell-to-cell communication (Hart et al., 1995; Szabo and Soltesz, 2015), and a *space code* in the hippocampus (O'Keefe and Burgess, 1996, 2005; Hafting et al., 2005; Brandon and Hasselmo, 2009; Papoutsi et al., 2009). It may be worth recalling that John O'Keefe, May-Britt Moser and Edvard Moser won the Nobel Prize for Medicine in 2014 precisely for the discovery that the cells of the hippocampus use the rules of a unique space code to build an internal map of the environment.

In the intermediate brain, in other words, the processing of neural signals takes place according to rules that are the neural equivalent of the signal transduction and the signal integration codes of the cell.

In addition to the organic world and to the neural world, furthermore, a third world of objects came into being. A world made of bows and arrows, of tables and chairs, of houses and cathedrals, of books and computers, of ships and space stations. It is our familiar world of culture, and in this case we know that it is based on codes because that is how we made it. The Morse code, the highway code, the articles of law, the rules of chess, the theorems of mathematics, the fairy tales and the value of money are all human conventions that are based on arbitrary rules and make up much of the universe in which we live.

There are, in conclusion, three types of objects in life – organic, neural and cultural objects – and they were the result of distinct types of evolution. For a long time it has been thought that codes exist only in the human world, but now we realize that they have been present at all levels in life, from the molecular level of the genetic code to the higher levels of the neural codes and of the codes of culture.

10. The definition of code

A code has been defined as "*a mapping between the objects of two independent worlds that is implemented by the objects of a third world called adaptors*" (Barbieri, 2003).

The genetic code, for example, is a mapping between codons and amino acids that is implemented by transfer-RNAs; a signal-transduction code is a mapping between first and second messengers that is implemented by molecular receptors; the Morse code is a mapping between the letters of the alphabet and groups of dots-and-dashes that is implemented by neural circuits according to a set of learned rules.

We have in this way a clear definition of code, but it must be underlined that it is by no means the only one and that other definitions have been adopted. In communication engineering, in particular, the transmission of messages is submitted to anti-noise operations, known as *source codes*, *channel codes* and *error-correcting codes*, whose purpose is to spot the errors that inevitably take place in real situations. One of these codes, for example, is the rule that every letter is repeated three times; in this way, the word *abc* is replaced by *aaabbbccc* and this allows the machine to recognize (and correct) a variety of errors. The loss of a letter, for example, would transform a triplet (*aaa*) into a duplet (*aa*) whereas the change of a letter would transform a uniform triplet (*aaa*) into a non-uniform one (*xaa*, *axa*, or *aax*).

The key point is that the engineering codes transform a linear sequence (*abc*) into another linear sequence (*aaabbbccc*) that belongs to *the same alphabet world*, whereas the genetic code, for example, transforms a linear sequence of nucleotides into a linear sequence of amino acids that belong to *a different alphabet world*. The result is that the genetic code and the other organic codes require *adaptors* to create bridges between two worlds, whereas no such adaptors are required in the engineering codes.

The actual situation, however, is more complicated than that, because *codes-without-adaptors* have been reported in biology too. In the 1980s and 90s, Edward Trifonov proposed that the genome carries several codes simultaneously, in addition to the classic genetic code, and gave them the collective name of *sequence codes* (Trifonov, 1989, 1996, 1999). This conclusion is based on Trifonov's definition that *"a code is any pattern in a sequence which corresponds to one or another specific biological function"* (Trifonov, 1996, p. 424). As in the case of the engineering codes, the sequence codes transform sequences into other sequence that belong *to the same alphabet world*, and no adaptors are required.

In addition to the codes described by Trifonov, furthermore, various other sequence codes have been brought to light. Among them, the *transcription regulatory codes* (Stergachis et al., 2013; Weatheritt and Babu, 2013), the *translation regulation codes* (Kindler et al., 2005), the *operon c*odes (Salgado et al., 2000; Blumenthal, 2004), the riboswitch codes (Nudler and Mironov, 2004; Vitreschak et al., 2004; Tucker and Breaker, 2005) and the histidine kinase code (Bilwes et al., 1999).

We must acknowledge, in other words, that codes can be defined in different ways, and we should not be surprised by this because the same has happened to other fundamental concepts in science.

The concept of *information*, for example, in biology has been defined as the linear order of nucleotides in a gene, whereas in engineering it has been defined as the *probability* of extracting a sequence from a pool of sequences and is expressed by the entropy-like formula introduced by Claude Shannon (1948). In reality, there is no contradiction between these two definitions of information because they represent two distinct aspects of reality, and the same is probably true for the different definitions of code.

What is essential in any code is the *arbitrariness* of its rules, and we may well acknowledge that there are arbitrary rules of correspondence between objects that belong to the same world as well as between objects that belong to two different worlds.

11. The concept of artifact-making

The defining feature of any code is its *arbitrariness*, the fact that its rules are not determined by the laws of physics and chemistry. But how can arbitrary rules exist in the molecular world? How could they have come into being? The Stereochemical theory, as we have seen, claimed that the genetic code is ultimately a *metaphor* because its rules were determined by chemistry. Günther Wächtershäuser expressed this concept in no uncertain terms: "*. . .If we could ever trace the historic process backwards far enough in time, we would wind up with an origin of life in purely chemical processes*" (Wächtershäuser 1997, p. 483).

This conclusion is based on the idea that all natural processes can be described in terms of physical quantities such as space, time, mass, energy, etc. . ., but in reality this is strictly true only in processes that occur spontaneously, whereas genes and proteins are *not* produced by spontaneous operations in living systems. They are produced by molecular machines that physically stick their subunits together and are therefore *manufactured molecules*, *molecular artefacts*. This in turn means that all biological structures are manufactured, not spontaneous, and therefore that the whole of life is *artefact-making* (Barbieri, 2006).

Spontaneous genes and spontaneous proteins did appear on the primitive Earth but did not evolve into the first cells because they did not have biological *specificity*. They gave origin instead to *molecular machines* and it was these machines and their products that evolved into the first cells.

The simplest molecular machines that appeared spontaneously on the primitive Earth were molecules that could stick other molecules together, first at random (*bondmakers*) and then in

the order provided by a template (*copymakers*). These molecules started manufacturing polymers such as polypeptides, polynucleotides and polysaccharides, and had the potential to produce them indefinitely thus increasing dramatically their presence in the world.

Biochemistry has long argued that all organic reactions are catalyzed processes, but molecular biology has discovered that genes and proteins require not only catalysts but also *templates*. The catalysts join the subunits together, and the templates provide the *order* in which the subunits are assembled. This is precisely the characteristic that divides spontaneous molecules from manufactured ones. In spontaneous molecules the order of the components comes *from within* the molecules, from *internal* factors, whereas in genes and proteins it comes *from without*, from *external* templates.

The description of biological processes, on the other hand, does not require only space, time, mass, energy, etc., because sequences and coding rules are absolutely essential to that purpose. We simply cannot describe the transmission of genes or the synthesis of proteins without their sequences, and we cannot replace these sequences with anything else, which means that using sequences and coding rules to describe living systems is perfectly equivalent to using space, time, mass and energy to describe physical systems (Barbieri, 2015a). Sequences and coding rules, in other words, are true *observables*, like the quantities of physics and chemistry, and are *a sine qua non* to the scientific study of life.

What is particularly important, to our purposes, is that the concept of artefact-making explains how it is possible that life evolved from inanimate matter and yet it is fundamentally different from it. Life evolved from inanimate matter because it was chemical evolution that gave origin to the first molecular machines; life is fundamentally different from inanimate matter because the molecular machines gave origin to the new world of molecular artefacts. The divide between matter and life is real because *matter is made of spontaneous objects* whereas *life is made of manufactured objects*.

12. The concept of codepoiesis

Modern biology describes the cell as an *autopoietic* system, a *system that fabricates itself* (Maturana and Varela, 1980; Hofmeyr, 2007). The concept of autopoiesis, or self-fabrication, describes the ability of living systems to produce their own components and eventually to produce copies of themselves. It must be underlined, however, that this is not always what takes place in life. Before the origin of the genetic code, for example, specific proteins could not exist, and the ancestral systems were producing descendants that were inevitably *different* from themselves. They were not autopoietic systems. Autopoiesis, in other words, did not exist before the first cells, so it was not the mechanism that gave origin to them.

Before the genetic code, the ribonucleoprotein apparatus of the common ancestor was engaged in the process of evolving coding rules and was therefore a *code exploring system*. After the origin of the genetic code, on the other hand, the situation completely changed. No other modification in the coding rules was accepted and the apparatus became a *code conservation system*. Another part of the system, however, maintained the potential to evolve other codes and behaved as a new *code exploring system*. In the ancestors of the eukaryotes, for example, the cells had *a code conservation part* for the genetic code, but also a *code exploring part* for the splicing codes and for the histone code.

This tells us something important about life. The origin of the first cells was based on the ability of the ancestral systems to *generate* the rules of the genetic code, and the subsequent evolution of the cells was based on two complementary processes: one was the *generation* of new organic codes and the other was the *conservation* of the existing ones. Taken together, these two processes are the two sides of a biological phenomenon that has been referred to as *codepoiesis* (Barbieri, 2012).

In modern biology, the conservation of the genetic code is explained by saying that any code is a set of 'constraints' (Crick, 1968), an idea suggested by the distinction that exists in physics between the *laws* and *constraints*. All planets, for example, are formed according to the universal laws of physics, and yet they are all different because their individual features were determined by the local constraints that affected their historical development.

The conclusion that the genetic code is a set of constraints is formally correct because it consists in a small number of rules selected from a large number of possibilities and represents therefore a limitation of the potentialities. It must be underlined, however, that the rules of the genetic code are *biological* constraints, not *physical* ones. They are *biologically generated* rules and in no way can be assimilated to physical constraints. This is because the genes of the genetic code are constantly subject, like all other genes, to mutation and neutral drift. They are in a continuous state of flux and the fact that they have been highly conserved in evolution means that there is a biological mechanism that *actively and continuously* restores their original structure. The conservation of the genetic code, in other words, is not the passive result of physical constraints that somehow 'freeze' it. It can only be the result of an active biological mechanism that is continuously at work, the mechanism that has been referred to as *codepoiesis*.

The ancestral systems that gave origin to the first cells were not autopoietic systems but they had to be codepoietic systems, and this tells us that the deep logic of life is codepoiesis, not autopoiesis.

13. The concept of cell organization

The greatest generalization of biology is the cell theory, the idea that cells always derive from preexisting cells, and this means that the information of the genes can only be expressed in the context provided by a cell. What must be transmitted from one generation to the next, in other words, is not only the genes but also the cellular context, or *cell organization*, that allows their expression (Harold, 2014).

Unfortunately, the concept of cell organization has been extremely elusive and no general agreement has yet been reached. It has been pointed out that the cell membrane must have a key role because it has the remarkable property that it is never constructed *de novo*. Membranes are always grown from pre-existing membranes, and this has led to the concept of *membrane heredity*, the idea that membranes are passed down from one generation to the next in an uninterrupted chain of descent (Blobel, 1980; Sapp, 1987; Cavalier-Smith, 2000; Harold, 2005).

Many supramolecular structures of the cell are produced by self-assembly from their components, but two of them are produced by growth from pre-existing structures. Chromosomes are produced from pre-existing chromosomes and membranes from pre-existing membranes, and they carry two very different types of information: chromosomes transmit genetic instructions, whereas membranes transmit architectural order. This makes it likely that membranes play a role in the transmission of organization, but they seem to account only for some aspects of it, not for the for the whole of cell organization.

A new approach to this problem has come from the discovery that the genetic code has been highly conserved in evolution. The long term conservation of the organic codes implies that they are transmitted from one generation to the next more faithfully than all other cell components, and this means that *the highest priority of a cell is to transmit its codes intact to the next generation; everything else is not so strictly conserved and can change within much wider limits*.

Cell organization, on the other hand, is associated with a pattern of relationships that is faithfully transmitted to the descendants, and this amounts to saying that the organization of a cell is *primarily* due to the network of its organic codes. There may well be other components that contribute to cell organization, but the role of the organic codes is bound to be crucial because it is the codes that create a long lasting network of relationships between the molecules of the cell.

We may even be tempted to consider a new model of the cell that takes into account only the components that are highly conserved in evolution, and which can be expressed in this way: "*the cell is a codepoietic system, i.e., a system that is capable of generating and conserving its own codes*" (Barbieri, 2015a).

14. A biology with codes

The existence of many organic codes in life is first and foremost *an experimental fact* – let us never forget this – but also more than that. It is one of those facts that have extraordinary theoretical consequences.

The first is the idea that the organic codes give us a new solution to the problem of complexity. More precisely, the fact that the eukaryotes have many more organic codes than prokaryotes suggests that they became more complex precisely because they evolved more organic codes. Which is equivalent to saying that the complexity of a cell is primarily due to the network of its organic codes.

The second theoretical consequence is an entirely new view of macroevolution. A code creates arbitrary relationships and in so doing it brings *absolute novelties* into existence because it gives origin to objects that did not exist before in the universe. Statistical proteins, for example, could appear spontaneously on the primitive Earth and elsewhere, but *specific proteins* came into being only after the origin of the genetic code. More than that. The evolution of the individual rules of a code can take an extremely long time, but the origin of a new organic code corresponds to the appearance of a complete set of rules and from a geological point of view that amounts to *a sudden* event. The great discontinuities of the history of life, in other words, can be explained as the result of the appearance of new organic codes, i.e., of new absolute novelties in the living world.

The third theoretical consequence is the place that the organic codes have in life. It may be argued, for example, that genetic code is only one component of the apparatus of protein synthesis, and it is the evolution of this whole apparatus that counts, not the evolution of a small part of it. Let us take a look therefore at the evolution of that larger apparatus. What we actually see is that the apparatus of protein synthesis evolved in many different ways in different lineages; the ribosomes, for example, have average molecular weights of 2 millions in prokaryotes and 4 millions in eukaryotes, and their actual molecular weights change from species to species; the eukaryotes have many more ribosomal proteins and much heavier ribosomal RNAs than prokaryotes, and many details of protein synthesis change even between Archaea and Bacteria.

What comes out, in the end, is that the only thing that all apparatuses of protein synthesis have in common is the genetic code, and this tells us something important. It tells us that the rest of the apparatus could change in countless different ways, but not its code. It tells us that the rules of the organic codes are the great invariants of life, the sole entities that have been conserved for billions of years while everything else has been changed.

The organic codes, in conclusion, are not just another novelty that must be added to the countless other characteristics of living systems. They are the most conserved components of life, and their study – the new research field of Code Biology – is destined to become an increasingly relevant part of the life sciences.

References

Acevedo-Rocha, C., Budisa, N., 2016. Xenomicrobiology: a roadmap for genetic code engineering. Microb. Biotechnol. 9, 666–676.

Adl, S.M., Simpson, A.B.G., Farmer, M.A., Andersen, R.A., Anderson, O.R., Barta, J.R., et al., 2005. The new higher-level classification of eukaryotes with emphasis on the taxonomy of protists. J. Eukaryot. Microbiol. 52, 399–451.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2007. Molecular Biology of the Cell, 5th ed. Garland, New York.

Baldauf, S.L., 2003. The deep roots of eukaryotes. Science 300, 1703–1706.

Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencow, B.J., Frey, B.J., 2010. Deciphering the splicing code. Nature 465, 53–59.

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

Barbieri, M., 2006. Life and Semiosis: the real nature of information and meaning. Semiotica 158, 233–254.

Barbieri, M., 2012. Codepoiesis – the deep logic of life. Biosemiotics 5 (3), 297–299.

Barbieri, M., 2014. From biosemiotics to code biology. Biol. Theory 9 (2), 239–249.

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

Barbieri, M., 2015b. Evolution of the genetic code: the ribosome-oriented model. Biol. Theory 10 (4), 301–310.

Barbieri, M., 2016. From the common ancestor to the first cells: the code theory. Biol. Theory 11, 102–112.

Barbieri, M., 2017. How did the eukaryotes evolve? Biol. Theory 12, 13–26.

Barghoorn, E.S., Tyler, S.M., 1965. Microfossils from the Gunflint chert. Science 147, 563–577.

Barisic, M., Maiato, H., 2016. The tubulin code: a navigation system for chromosomes during mitosis. Trends Cell Biol. 26 (10), 766–775.

Basañez, G., Hardwick, J.M., 2008. Unravelling the Bcl-2 apoptosis code with a simple model system. PLoS Biol. 6 (6), e154, http://dx.doi.org/10.1371/journal.pbio.0060154.

Berger, S.L., 2007. The complex language of chromatin regulation during transcription. Nature 447, 407–412.

Bilwes, A.M., Alex, L.A., Crane, B.R., Simon, M.I., 1999. Structure of CheA, a signal-transducing histidine kinase. Cell 96, 131–141.

Blobel, G., 1980. Intracellular membrane topogenesis. Proc. Natl. Acad. Sci. U. S. A. 77, 1496–1500.

Blumenthal, T., 2004. Operons in eukaryotes. Brief. Funct. Genomics Proteomics 3 (3), 199–211.

Bollenbach, T., Kalin Vetsigian, K., Kishony, R., 2007. Evolution and multilevel optimization of the genetic code. Genome Res. 17, 401–404.

Brandon, M.P., Hasselmo, M.E., 2009. Sources of the spatial code within the hippocampus. Biol. Rep. 1, 3–7.

Buckeridge, M.S., De Souza, A.P., 2014. Breaking the Glycomic Code of cell wall polysaccharides may improve second-generation bioenergy production from biomass. Bioenergy Res. 7, 1065–1073.

Budisa, N., 2004. Prolegomena to future efforts on genetic code engineering by expanding its amino acid repertoire. Angew. Chem. Int. Ed. 43, 3387–3428.

Buratti, E., Baralle, M., Baralle, F.E., 2006. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. Nucleic Acids Res. 34, 3494–3510.

Cavalier-Smith, T., 2000. Membrane heredity and early chloroplast evolution. Trends Plant Sci. 5, 174–182.

Chakraborti, S., Natarajan, K., Curiel, J., Janke, C., Liu, J., 2016. The emerging role of the tubulin code: from the tubulin molecule to neuronal function and disease. Cytoskeleton 73, 9, http://dx.doi.org/10.1002/cm.21290.

Cooper, T.A., Wan, L., Dreyfuss, G., 2009. RNA and disease. Cell 136, 777–793.

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

De Beule, J., Hovig, E., Benson, M., 2011. Introducing dynamics into the field of biosemiotics. Biosemiotics 4 (1), 5–24.

De Beule, J., 2014. Sketch for a theory of evolution based on coding. Biosemiotics 7 (2), 181–201.

Dhir, A., Buratti, E., van Santen, M.A., Lührmann, R., Baralle, F.E., 2010. The intronic splicing code: multiple factors involved in ATM pseudoexon definition. EMBO J. 29, 749–760.

Di Lorenzo, P.M., 2000. The neural code for taste in the brain stem: response profiles. Physiol. Behav. 69, 87–96.

Dudai, Y., 1999. The smell of representations. Neuron 23, 633–635.

Füllgrabe, J., Hajji, N., Joseph, B., 2010. Cracking the death code: apoptosis-related histone modifications. Cell Death Differ. 17, 1238–1243.

Farina, A., Pieretti, N., 2014. Acoustic codes in action in a soundscape context. Biosemiotics 7 (2), 321–328.

Fitch, W.M., Upper, K., 1987. The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. Cold Spring Harb. Symp. Quant. Biol. 52, 759–767.

Flames, N., Pla, R., Gelman, D.M., Rubenstein, J.L.R., Puelles, L., Marìn, O., 2007. Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. J. Neurosci. 27 (36), 9682–9695.

Fox, G.E., 2010. Origin and evolution of the ribosome. Cold Spring Harb. Perspect. Biol. 2, a003483.

Freeland, S.J., Hurst, L.D., 1998. The genetic code is one in a million. J. Mol. Evol. 47, 238–248.

Fu, X.D., 2004. Towards a splicing code. Cell 119, 736–738.

Görlich, D., Dittrich, P., 2013. Molecular codes in biological and chemical reaction networks. PLoS One 8 (1), e54, http://dx.doi.org/10.1371/journal.pone.0054694 (694).

Görlich, D., Artmann, S., Dittrich, P., 2011. Cells as semantic systems. Biochim. Biophys. Acta 1810 (10), 914–923.
Gabius, H.-J., 2000. Biological information transfer beyond the genetic code: the sugar code. Naturwissenschaften 87, 108–121.
Gabius, H.-J., 2009. The Sugar Code. Fundamentals of Glycosciences. Wiley-Blackwell.
Gamow, G., 1954. Possible relation between deoxyribonucleic acid and protein structures. Nature 173, 318.
Gilis, D., Massar, S., Cerf, N.J., Rooman, M., 2001. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. Genome Biol. 2, 41–49.
Gimona, M., 2008. Protein linguistics and the modular code of the cytoskeleton. In: Barbieri, M. (Ed.), The Codes of Life: The Rules of Macroevolution. Springer, Dordrecht, pp. 189–206.
Hafting, T., Fyhn, M., Molden, S., Moser, M.B., Moser, E.I., 2005. Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801–806.
Haig, D., Hurst, L.D., 1991. A quantitative measure of error minimization in the genetic code. J. Mol. Evol. 33, 412–417.
Hallock, R.M., Di Lorenzo, P.M., 2006. Temporal coding in the gustatory system. Neurosci. Behav. Rev. 30, 1145–1160.
Harold, F.M., 2005. Molecules into cells: specifying spatial architecture. Microbiol. Mol. Biol. Rev. 69, 544–564.
Harold, F.M., 2014. In Search of Cell History. The Evolution of Life's Building Blocks. The University of Chicago Press, Chicago and London.
Hart, A.C., Sims, S., Kaplan, J.M., 1995. Synaptic code for sensory modalities revealed by C: elegans GLR-1 glutamate receptor. Nature 378, 82–85.
Hartman, M.C.T., Josephson, K., Lin, C.-W., Szostak, J.W., 2007. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. PLoS One 2 (10), e972, http://dx.doi.org/10.1371/journal.pone.0000972.
Hofmeyr, J.-H.S., 2007. The biochemical factory that autonomously fabricates itself: a systems-biological view of the living cell. In: Boogard, F., Bruggeman, F., Hofmeyr, J., Westerhoff, H. (Eds.), Towards a Philosophy of Systems Biology. Elsevier, pp. 217–242 (Chapter 10).
Hou, Y.-M., Schimmel, P., 1988. A simple structural feature is a major determinant of the identity of a transfer RNA. Nature 333, 140–145.
Hunt, P., Whiting, J., Nonchev, S., Sham, M.-H., Marshall, H., Graham, A., Cook, M., Alleman, R., Rigby, P.W., Gulisano, M., 1991. The branchial *Hox* code and its implications for gene regulation, patterning of the nervous system and head evolution. Development 2, 63–77.
Janke, C., 2014. The tubulin code: molecular components, readout mechanisms, and functions. J. Cell Biol. 206 (4), 461–472.
Jessell, T.M., 2000. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nat. Genet. 1, 20–29.
Jun, S.-R., Sims, G.E., Wu, G.A., Kim, S.H., 2010. Whole genome phylogeny of prokaryotes by feature frequency profiles: an alignment-free method with optimal resolution. Proc. Natl. Acad. Sci. U. S. A. 107, 133–138.
Kühn, S., Hofmeyr, J.-H.S., 2014. Is the Histone Code an organic code? Biosemiotics 7 (2), 203–222.
Keeling, P.J., Burger, G., Durnford, D.G., Lang, B.F., Lee, R.W., et al., 2005. The tree of the eukaryotes. Trends Ecol. Evol. 20, 670–676.
Kessel, M., Gruss, P., 1991. Homeotic tansformation of murine vertebrae and concomitant alteration of hox codes induced by retinoic acid. Cell 67, 89–104.
Kindler, S., Wang, H., Richter, D., Tiedge, H., 2005. RNA transport and local control of translation. Annu. Rev. Cell Dev. Biol. 21, 223–245.
Knoll, A.H., 2003. Life on a Young Planet. The First Three Billion Years of Evolution on Earth. Princeton University Press, Princeton NJ.
Komander, D., Rape, M., 2012. The ubiquitin code. Annu. Rev. Biochem 81, 203–229.
Koonin, E.V., Novozhilov, A.S., 2009. Origin and evolution of the genetic code: the universal enigma. IUBMB Life 61 (2), 99–111.
Koonin, E.V., 2003. Comparative genomics, minimal gene-Sets and the last universal common ancestor. Nat. Rev. Microbiol. 1, 127–136.
Koonin, E.V., 2012. The Logic of Chance. The Nature and Origin of Biological Evolution. Pearson Education, Upper Saddle River, New Jersey.
Lane, N., 2015. The Vital Question. Energy, Evolution and the Origins of Complex Life. WW Norton and Company.
Levin, M., 2014. Endogenous bioelectrical networks store non-genetic patterning information during development and regeneration. J. Physiol. (Lond.) 592, 2295–2305.
Ling, J., O'Donoghue, P., Söll, D., 2015. Genetic code flexibility in microorganisms: novel mechanisms and impact on bacterial physiology. Nat. Rev. Microbiol. 13, 707–721.
Maizels, N., Weiner, A.M., 1987. Peptide-specific ribosomes, genomic tags and the origin of the genetic code. Cold Spring Harb. Symp. Quant. Biol. 52, 743–749.
Maraldi, N.M., 2008. A lipid-based code in nuclear signalling. In: Barbieri, M. (Ed.), The Codes of Life: The Rules of Macroevolution. Springer, Dordrecht, pp. 207–221.
Marquard, T., Pfaff, S.L., 2001. Cracking the transcriptional code for cell specification in the neural tube. Cell 106, 651–654.
Matlin, A., Clark, F., Smith, C., 2005. Understanding alternative splicing: towards a cellular code. Nat. Rev. Mol. Cell Biol. 6, 386–398.
Maturana, H.R., Varela, F.J., 1980. Autopoiesis and Cognition: The Realisation of the Living. D. Reidel Publishing Company, Dordrecht, Holland.
Nicolelis, M., Ribeiro, S., 2006. Seeking the neural code. Sci. Am. 295, 70–77.
Nicolelis, M., 2011. Beyond Boundaries: The New Neuroscience of Connecting Brains with Machines and How It Will Change Our Lives. Times Books, New York.
Nitta, I., Kamada, Y., Noda, H., Ueda, T., Watanabe, K., 1998. Reconstitution of peptide bond formation. Science 281, 666–669.
Nudler, E., Mironov, A.S., 2004. The riboswitch control of bacterial metabolism. Trends Biochem. Sci. 29 (1), 11–17.
O'Keefe, J., Burgess, N., 1996. Geometric determinants of the place fields of hippocampal neurons. Nature 381, 425–428.
O'Keefe, J., Burgess, N., 2005. Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. Hippocampus 15, 853–866.
Osawa, S., 1995. Evolution of the Genetic Code. Oxford University Press, New York.
Papoutsi, M., de Zwart, J.A., Jansma, J.M., Pickering, M.J., Bednar, J.A., Horwitz, B., 2009. From phonemes to articulatory codes: an fMRI study of the role of broca's area in speech production. Cereb. Cortex 19, 2156–2165.
Pertea, M., Mount, S.M., Salzberg, S.L., 2007. A computational survey of candidate exonic splicing enhancer motifs in the model plant Arabidopsis thaliana. BMC Bioinf. 8, 159.
Raunser, S., Gatsoiannis, C., 2015. Deciphering the tubulin code. Cell 161 (5), 960–961.
Ray, A., van der Goes van Naters, W., Shiraiwa, T., Carlson, J.R., 2006. Mechanisms of odor receptor gene choice in drosophila. Neuron 53, 353–369.
Redies, C., Takeichi, M., 1996. Cadherine in the developing central nervous system: an adhesive code for segmental and functional subdivisions. Dev. Biol. 180, 413–423.
Ruiz i Altaba, A., Nguien, V., Palma, V., 2003. The emergent design of the neural tube: prepattern, SHH morphogen and GLI code. Curr. Opin. Genet. Dev. 13, 513–521.
Salgado, H., Moreno-Hagelsieb, G., Smith, T.F., Collado-Vides, J., 2000. Operons in Escherichia coli: genomic analyses and predictions. Proc. Natl. Acad. Sci. U. S. A. 97 (12), 6652–6657.
Sapp, J., 1987. Beyond the Gene: Cytoplasmic Inheritance and the Struggle for Authority in Genetics. Oxford University Press, Oxford UK.
Schimmel, P., 1987. Aminoacyl tRNA synthetases: general scheme of structure-function relationship in the polypeptides and recognition of tRNAs. Ann. Rev. Biochem. 56, 125–158.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–424 (623–656).
Shapiro, L., Colman, D.R., 1999. The diversity of cadherins and implications for a synaptic adhesive code in the CNS. Neuron 23, 427–430.
Simonson, A.B., Servin, J.A., Skophammer, R.G., Herbold, C.W., Rivera, M.C., Lake, J.A., 2005. Decoding the genomic tree of life. Proc. Natl. Acad. Sci. U. S. A. 102, 6608–6613.
Snel, B., Huynen, M.A., Dulith, B.A., 2005. Genome trees and the nature of genome evolution. Ann. Rev. Microbiol. 59, 191–209.
Solis, A.S., Shariat, N., Patton, J.G., 2008. Splicing fidelity, enhancers, and disease. Front. Biosci. 13, 1926–1942.
Stergachis, A.B., Haugen, E., Shafer, A., Fu, W., Vernot, B., Reynolds, A., Raubitschek, A., Ziegler, S., LeProust, E.M., Akey, J.M., Stamatoyannopoulos, J.A., 2013. Exonic transcription factor binding directs codon choice and affects protein evolution. Science 342, 1367–1372.
Strahl, B.D., Allis, D., 2000. The language of covalent histone modifications. Nature 403, 41–45.
Szabo, G.G., Soltesz, I., 2015. Pass-through code of synaptic integration. Neuron 87, 1124–1126.
Tazi, J., Bakkour, N., Stamm, S., 2009. Alternative splicing and disease. Biochim. Biophys. Acta 1792, 14–26.
Tomkins, M.G., 1975. The metabolic code. Science 189, 760–763.
Trifonov, E.N., 1989. The multiple codes of nucleotide sequences. Bull. Math. Biol. 51, 417–432.
Trifonov, E.N., 1996. Interfering contexts of regulatory sequence elements. Cabios 12, 423–429.
Trifonov, E.N., 1999. Elucidating sequence codes: three codes for evolution. Ann. N. Y. Acad. Sci. 870, 330–338.
Tseng, A.S., Levin, M., 2013. Cracking the bioelectric code. Probing endogenous ionic controls of pattern formation. Commun. Integr. Biol. 6 (1), 1–8.
Tucker, B.J., Breaker, R.R., 2005. Riboswitches as versatile gene control elements. Curr. Opin. Struct. Biol. 15 (3), 342–348.
Turner, B.M., 2000. Histone acetylation and an epigenetic code. Bioessays 22, 836–845.
Turner, B.M., 2002. Cellular memory and the histone code. Cell 111, 285–291.
Turner, B.M., 2007. Defining an epigenetic code. Nat. Cell Biol. 9, 2–6.
Verhey, K.J., Gaertig, J., 2007. The tubulin code. ABBV Cell Cycle 6 (17), 2152–2160.
Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., Gelfand, M.S., 2004. Riboswitches: the oldest mechanism for the regulation of gene expression? Trends Genet. 20 (1), 44–50.
Wächtershäuser, G., 1997. The origin of life and its methodological challenge. J. Theor. Biol. 187, 483–494.
Wang, Z., Burge, C., 2008. Splicing regulation: from a part list of regulatory elements to an integrated splicing code. RNA 14, 802–813.
Wang, G.S., Cooper, T.A., 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat. Rev. Genet. 8, 749–761.
Weatheritt, R.J., Babu, M.M., 2013. The hidden codes that shape protein evolution. Science 342, 1325–1326.
Woese, C.R., Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. U. S. A. 74, 5088–5090.

Woese, C.R., Dugre, D.H., Dugre, S.A., Kondo, M., Saxinger, W.C., 1966. On the fundamental nature and evolution of the genetic code. Cold Spring Harb. Symp. Quant. Biol. 31, 723–736.
Woese, C.R., Kandler, O., Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eukarya. Proc. Natl. Acad. Sci. U. S. A. 87, 4576–4579.
Woese, C.R., 1965. Order in the genetic code. Proc. Natl. Acad. Sci. U. S. A. 54, 71–75.
Woese, C.R., 2000. Interpreting the universal phylogenetic tree. Proc. Natl. Acad. Sci. U. S. A. 97, 8392–8396.
Wolf, Y.I., Koonin, E.V., 2007. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. Biol. Direct 2, 14.
Wong, J.T., 1975. A co-evolution theory of the genetic code. Proc. Natl. Acad. Sci. U. S. A 72 (5), 1909–1912.
Wong, J.T., 1981. Coevolution of genetic code and amino acid biosynthesis. Trends Biochem. Sci. 6, 33–36.
Zuckerkandl, E., Pauling, L., 1965. Molecules as documents of evolutionary history. J. Theor. Biol. 8, 357–366.