Biol Theory (2017) 12:13–26
DOI 10.1007/s13752-016-0253-3

ORIGINAL ARTICLE

How Did the Eukaryotes Evolve?

Marcello Barbieri[1]

Received: 15 July 2016 / Accepted: 15 September 2016 / Published online: 19 October 2016

Abstract The fossil record shows that the stromatolites built by cyanobacteria 2 and 3 billion years ago are virtually identical to those built by their modern descendants, which is just a part of much evidence revealing that bacteria have barely changed in billions of years. They appeared very early in the history of life and have conserved their complexity (in terms of size, shape, and number of components) ever since. The eukaryotes, however, did the opposite. They repeatedly increased the complexity of their cells and eventually broke the cellular barrier and gave origin to all living creatures that we see around us. This gives us one of the major problems in evolution: why have the prokaryotes maintained the same complexity throughout the history of life while the eukaryotes have become increasingly more complex? Here it is shown that a solution does exist, but it is based on experimental data that so far have largely been ignored. It is based on the discovery that, in addition to the genetic code, many other organic codes exist in living systems. The potential to generate organic codes was already present in the common ancestor but was not transmitted indefinitely to all its descendants. After the genetic code and the signal transduction codes that gave origin to the first cells, the prokaryotes evolved no other organic code, whereas the ancestors of the eukaryotes continued to explore the coding space and gave origin to splicing codes, histone code, cytoskeleton codes, tubulin code, compartment codes, and sequence codes. This experimental fact 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 codes into existence.

Keywords Common ancestor · Complexity · Eukaryotes · Organic codes · Primary kingdoms · Prokaryotes · Sequence codes

The Problems of Cellular Evolution

One of the greatest discoveries of paleontology is that our planet was inhabited exclusively by free-living cells, or microorganisms, for the first 3 billion years of the history of life (Schopf 1999). It is an experimental fact, as well, that cells are either prokaryotes (without a nucleus) or eukaryotes (well nucleated), and we have therefore the problem of understanding how those two types of microorganisms evolved during those billions of years.

The fossil record has revealed the presence of fossilized bacteria in most Precambrian rocks, and has shown that the stromatolites built by cyanobacteria 2 and 3 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. They have exploited virtually all sources of energy on Earth and have adapted to nearly all environments, but the extraordinary achievements of their metabolic versatility were obtained without changing the overall complexity of the bacterial cell. This point has been beautifully illustrated by 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

✉ Marcello Barbieri
brr@unife.it

[1] Department of Morphology and Embryology, University of Ferrara, Ferrara, Italy

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 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 maintained the same complexity through the history of life while the eukaryotes have become increasingly more complex?

Another great problem in evolution is the types of cells that started it.

The fact that all known cells contain a virtually universal genetic code implies that that code evolved in a population of primitive systems that has become known as the last universal common ancestor (LUCA). It must be underlined, on the other hand, that that ancestor did not transmit other universal features to its descendants. In particular it did not transmit a universal cell membrane. This means that it was the descendants of the common ancestor that evolved the modern cell membranes and gave origin to the first modern cells. But how did they do it? And how many types of cells descended from the common ancestor?

A major turning point in this field came in 1977, when Woese and Fox discovered that the phylogenetic trees obtained from ribosomal-RNAs divide all living creatures in three major groups: two types of bacteria that Woese and Fox (1977) called *archaebacteria* and *eubacteria*, and a third group representing the ancestors of the eukaryotic cells that they called *urkaryotes*.

This discovery has two outstanding implications:

(1) There are two fundamentally different types of bacteria in life (archaebacteria and eubacteria).
(2) The first ancestors of the eukaryotes were phylogenetically as old as the bacteria.

Later on, Woese renamed the three groups and proposed that all living systems belong to three distinct primary kingdoms, or domains, that were called *Archaea*, *Bacteria,* and *Eucarya* (Woese 1987, 2000; Woese et al. 1990).

Woese's first conclusion, the idea that there is a fundamental dichotomy between Archaea and Bacteria, has been fully confirmed. In Bacteria, for example, the cell membrane contains phospholipids, whereas in Archaea it contains isoprenoid lipids. In Bacteria the cell wall is made of peptidoglycans, whereas in Archaea it is made of proteinaceous material. Bacteria move by flagella that obtain energy from ion currents, whereas Archaea move by totally different organelles that obtain energy from ATP (Harold 2014). What has been much more controversial, instead, is Woese's second conclusion, the idea that the descendants of the last common ancestor gave origin not only to Archaea and Bacteria but also to the first ancestors of the eukaryotes, the cells that he called *urkaryotes*.

The Primary Kingdoms

The universal tree reconstructed by Woese was obtained from ribosomal RNAs, but in principle it should also be recovered from proteins because they too contain phylogenetic information. When the techniques of molecular phylogeny were applied to proteins, however, the results turned out much more complex than expected. Some proteins confirmed the three domains obtained from the ribosomal RNAs, but other proteins led to different phylogenetic trees (Brown and Doolittle 1997).

The solution to this puzzle came from the discovery that bacteria are regularly swapping genetic material with the process of horizontal gene transfer (Miller 1998). The pattern of a tree is realized when genes are transmitted virtually unchanged from one generation to the next, i.e., when descent is vertical. When genes instead are swapped horizontally in every generation they become part of many branches simultaneously and the resulting pattern is no longer a tree but a web (Doolittle 1999; Doolittle and Bapteste 2007).

The phylogenetic record, in other words, has been heavily blurred by horizontal gene transfer, and the three primary kingdoms emerge clearly only from molecules that have largely avoided that process, i.e., from molecules—like the ribosomal RNAs—that have been highly conserved in evolution.

A more powerful approach to cell phylogeny was taken up when it became possible to study not only individual molecules but 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 five or six major groups that radiated from what has been called the last eukaryotic common ancestor (LECA) (Baldauf 2003; Adl et al. 2005; Keeling et al. 2005; Koonin 2012).

Another outstanding result of genome phylogeny was the discovery that the eukaryotic cells received genes from both Archaea and Bacteria. More precisely, they received about 20 % of their genes from Bacteria and about 10 % from Archaea, while the remaining 70 % are exclusively found in eukaryotes and are referred to as eukaryotic signature genes (Lane 2015). The problem is that this experimental fact can be interpreted in two completely different ways.

One hypothesis is that the common ancestor gave origin to three primary kingdoms, as postulated by Woese, and the eukaryotes received genes from all three lineages. The other hypothesis is that the common ancestor gave origin only to two primary kingdoms—Archaea and Bacteria—and the eukaryotes appeared at a later stage as a result of an extraordinary cell fusion between an archaeon and a bacterium (Martin and Müller 1998; Lòpez-Garcia and Moreira 1999; Rivera and Lake 2004; Koonin 2007). In this case the eukaryotic signature genes would have been the result of extensive modifications of the original prokaryotic genes. This hypothesis is highly popular today, but Woese's idea of three primary kingdoms is fully compatible with the evidence and cannot be ruled out (Kurland et al. 2006).

What is most important, to our purposes, is that the above hypotheses do not give an answer to the problem of complexity. None of them explains why the prokaryotes have remained substantially the same throughout the history of life while the eukaryotes have become increasingly more complex.

As a matter of fact, a solution to this problem does exist, but is based on experimental data that so far have largely been ignored. It is based on the discovery that, in addition to the genetic code, many other organic codes are at work in living systems.

The Road Not Taken

A wide variety of organic codes have come to light in living systems in addition to the genetic code. Among them, the sequence codes (Trifonov 1989, 1996, 1999), the histone code (Strahl and Allis 2000; Turner 2000, 2007; Kühn and Hofmeyr 2014), the sugar code (Gabius 2000, 2009), the signal transduction codes (Barbieri 2003), the compartment codes (Barbieri 2003), the splicing codes (Barbieri 2003; Fu 2004; Matlin et al. 2005; Pertea et al. 2007; Barash et al. 2010; Dhir et al. 2010), the cytoskeleton codes (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 signaling 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), and the ubiquitin code (Komander and Rape 2012).

The fact that many organic codes exist today implies not only that they came into being during the history of life but also that they contributed to that history. So far, however, this has not been taken into account and the theories that have been proposed on cellular evolution have not even mentioned the organic codes that appeared in living systems after the genetic code.

There are various reasons for that but one of them is particularly worth mentioning. The very first model of the genetic code was the stereochemical theory, an idea proposed by Gamow in 1954 and later re-proposed by many other authors, which states that the relationships between codons and amino acids are determined by stereochemical affinities. This theory automatically implies that the genetic code is not a real code because its rules are the inevitable result of chemical processes and do not have the arbitrariness that is essential in any code.

It took a long time and much experimental work to overturn this conclusion. Eventually, however, it was shown that there are no deterministic links between codons and amino acids since any codon can be associated, in principle, to any amino acid (Schimmel 1987; Schimmel et al. 1993). Hou and Schimmel (1988), for example, introduced two extra nucleotides in a tRNA and found that that the resulting tRNA was carrying a different amino acid. Similar results have been obtained by many other modifications of the genetic code (Budisa 2004, 2014; Hartman et al. 2007; Ling et al. 2015; Acevedo-Rocha and Budisa 2016), thus proving that the number of possible connections between codons and amino acids is potentially unlimited. The genetic code, in other words, is a real code because it is a set of arbitrary rules that create a mapping between the objects of two independents worlds.

The stereochemical theory, on the other hand, has been the dominant model for many years in biology and its message that the genetic code is determined by chemistry is still hanging around. This probably explains the persistent belief that the organic codes are metaphorical entities, mere names that we use simply because they are intuitively appealing. Before discussing their role in evolution, therefore, we need to make sure that they are experimental realities, that there is nothing metaphorical about them.

Organic codes are relationships between two independent worlds of organic molecules and are necessarily implemented by a third type of molecules, called adaptors, that build a bridge between them. The adaptors are required because there is no necessary link between two independent worlds, and a fixed set of adaptors is required in order to guarantee the specificity of the mapping. The adaptors, in short, are the molecular fingerprints of the codes, and we can prove that an organic code exists if we find three things: (1) two independent worlds of molecules, (2) a potentially unlimited number of arbitrary connections between them implemented by adaptors, and (3) a selection of the adaptors (a set of coding rules) that ensures a specific mapping (Barbieri 2003).

It must be underlined that the term "code" has also been given to processes that are not based on adaptors, and this is clearly a potential source of confusion. In order to avoid misunderstandings, therefore, the term "organic code" is used here exclusively for the adaptor-based processes. With this premise, let us now take a look at some of the organic codes that have been discovered so far in addition to the genetic code.

The Signal Transduction Codes

Signal transduction is the process by which cells transform the signals from the environment, called first messengers, into internal signals, called second messengers. First and second messengers belong to two independent worlds because there are literally hundreds of first messengers (hormones, growth factors, neurotransmitters, etc.) but only four main families of second messengers (cyclic AMP, calcium ions, diacylglycerol, and inositol trisphosphate) (Alberts et al. 2007). The crucial point is that the molecules that perform signal transduction are true adaptors. They consist of three subunits: a receptor for the first messengers, an amplifier for the second messengers, and a mediator in between (Berridge 1985). This allows the transduction complex to perform two independent recognition processes, one for the first messenger and the other for the second messenger.

Laboratory experiments have proved that any first messenger can be associated with any second messenger, which means that there is a potentially unlimited number of arbitrary connections between them (Alberts et al. 2007). In signal transduction, in short, we find the three essential components of a code: (1) two independent worlds of objects (first messengers and second messengers), (2) a potentially unlimited number of arbitrary connections produced by adaptors, and (3) a set of coding rules (a selection of the adaptors) that ensures the specificity of the correspondence.

As for the role that the signal transduction codes had in evolution, a good clue comes from first examining the evolutionary role of the genetic code. The first genetic code that appeared on Earth was necessarily ambiguous because biological specificity had not yet come into existence (Fitch and Upper 1987; Osawa 1995). In that case, a sequence of codons was translated at some point into a protein and at some other time into a different protein, and the apparatus of protein synthesis was inevitably producing statistical proteins (Woese 1965). The evolution of the genetic code was therefore a process that steadily reduced and finally eliminated its ambiguity (Barbieri 2015a). When that happened, it became possible to translate genes into specific proteins and life as we know it—life based on biological specificity—came into existence.

The origin of a non-ambiguous genetic code, however, was not enough to produce 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 is the role that the signal transduction codes had 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 was these specific behaviors that marked 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 signaling systems (Harold 2014; Marijuan et al. 2015). This suggests that the descendants of the last common ancestor evolved along independent lines and gave origin to three types of cells by combining the universal genetic code with three distinct signal-transduction codes.

The Splicing Codes

All genes are copied into RNA molecules that are called primary transcripts, and all proteins are made from templates of RNA molecules that are called messenger RNAs. In bacteria, the primary transcripts are directly used as messenger RNAs, but in eukaryotes things are much more complex. The primary transcripts are first cut into pieces and then some of them (called introns) are removed, and the remaining pieces (called exons) are joined together to form a messenger RNA. These cutting-and-sealing operations are collectively known as splicing, and it has been shown that there are significant parallels with protein synthesis: (a) the splicing machines, known as spliceosomes, are huge molecular structures like ribosomes; (b) splicing employs small molecules, called small nuclear RNAs (snRNA), that are comparable to transfer-RNAs; and (c) the result of both processes is the assembly of molecules: splicing assembles messenger-RNAs from exons, whereas protein synthesis assembles proteins from amino acids.

The crucial point 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 an intron. In splicing, therefore, we find all the three essential components of a code: (1) two independent worlds of objects (primary transcripts and messenger RNAs), (2) a potentially unlimited number of arbitrary connections produced by adaptors, and (3) a set of coding rules that ensures the specificity of the correspondence. Splicing, in other words, is a codified process based on adaptors and takes place with sets of rules that have been referred to as splicing codes (Barbieri 2003; Fu 2004; Matlin et al. 2005; Wang and Burge 2008).

It must be underlined that there are two major complications in splicing with respect to protein synthesis. One is the fact that the order in which the exons are joined together can be shuffled in various ways, and this operation, called alternative splicing, allows the cell to obtain a whole family of proteins from the same gene. The expression of these proteins, furthermore, can change during embryonic development, which means that alternative splicing has a key role in the generation of biological diversity. It has been found, furthermore, that splicing mistakes have pathological effects that account for about one-fifth of all inherited diseases (Buratti et al. 2006; Wang and Cooper 2007; Solis et al. 2008; Cooper et al. 2009; Tazi et al. 2009).

The other great complication of splicing 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). 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.

The actual deciphering of the splicing codes is taking considerably longer than that of the genetic code because it is incredibly more complex. Let us keep in mind that the discovery of the genetic code has been facilitated by two particularly favorable features, more precisely, by the fact that (1) the adaptors are single molecules (the tRNAs) and (2) the coding units form a closed set (64 codons and 20 amino acids). In the case of splicing, instead, the adaptors are combinations of molecules (combinatorial codes), and the domain (or alphabet) of the codes is open and potentially unlimited.

The overall complexity of splicing is such that the most practical way of discovering its rules is by building computational models that are capable of predicting new splicing rules on the basis of existing data. Such models have already started appearing in the literature (Pertea et al. 2007; Barash et al. 2010; Dhir et al. 2010) and represent our first glimpse of the splicing codes.

The Histone Code

The classical double helix of the DNA has a width of two nanometers but in eukaryotes many segments of this filament are folded around groups of eight histone proteins and form blocks, called nucleosomes, that give the filament a beads-on-a-string appearance. This string, called chromatin, is almost six times thicker than the double helix and is further folded into spirals of nucleosome groups, called solenoids, that arrange it in fibers of increasing thickness and ultimately into the 600 nm fiber of the chromosome.

These multiple foldings allow the eukaryotic cells to pack their long chromosomes into the tiny space of their nuclei, and for this reason it was originally assumed that the histones have a purely packaging role. The experimental data, however, revealed that the "tails" of the histones (the parts that protrude from the surface of the nucleosomes) are subject to post-translational modifications (in particular acetylation, methylation, and phosphorylation) that have highly dynamic roles and are involved in the activation or in the repression of gene activity (Kornberg and Lorch 1999; Wu and Grunstein 2000).

The histone tails represent about 25–30 % of the histone mass, and their modifications can alter the chromatin either directly or indirectly. The direct modifications are those that physically open or close the molecular space (in particular the electrostatic barrier) that surrounds the genes and in this way control the transit of DNA-binding proteins (Hansen et al. 1998; Wolfe and Hayes 1999). Several discoveries, however, have shown that the most frequent effects are obtained by indirect mechanisms. In these cases, the modified histone tails provide "marks" on the surface of the nucleosomes that are recognized by specialized effector proteins that set in motion chains of biological reactions that eventually end in the activation or in the repression of specific genes (Agalioti et al. 2002; Peterson and Laniel 2004; Berger 2007; Gräff and Mansuy 2008).

A key breakthrough in this field came with the discovery that the histone modifications do not act individually, and the final result is due to a combination of histone marks rather than a single one. This led David Allis and colleagues to propose that the histone marks operate in combinatorial groups, like letters that are put together into the words of a molecular "language" and that was referred to as histone code (Strahl and Allis 2000; Jenuwein and Allis 2001).

The same concept was independently proposed by Turner (2000, 2002) who argued that there is an epigenetic code at the heart of the regulation mechanism based on histone tail modifications. He also noticed that there are both short-term and long-term effects in that mechanism. The short-term modifications change rapidly in response to external signals and represent a mechanism by which the genome quickly responds to the environment (Schreiber and Berstein 2002). The long-term modifications, instead, are those that are set up at early stages of embryonic development and allow the transcription of specific genes at more advanced stages (Turner 2007).

Kühn and Hofmeyer (2014) have pointed out that the effector proteins of the histone code have two distinct domains: one that recognizes histone modifications and one that initiates biological reactions. It has been shown, for example, that the acetylated lysines are specifically recognized only by the bromodomains of the effector proteins (Owen et al. 2000; Mujtaba et al. 2007). The methylated amino acids are recognized by a greater variety of domains, but again each recognition step is absolutely specific (Maurer-Stroh et al. 2003; Kim et al. 2006). The effector proteins, in other words, perform two independent recognition processes and behave therefore as true adaptors.

These days, in conclusion, a large number of data supports the idea that the regulation of genetic activity by histone modifications plays a fundamental role in all eukaryotes and is based on the rules of a combinatorial code that has become known as the histone code.

The Cytoskeleton Codes

A cytoskeleton is essential for typical eukaryotic processes such as phagocytosis, mitosis, meiosis, ameboid movement, organelle assembly, and three-dimensional organization of the cell, i.e., for all those features that make eukaryotic cells so radically different from bacteria. The actual cytoskeleton, in reality, is an integrated system of three different cytoskeletons made of filaments (microfilaments, microtubules, and intermediate filaments), each of which gives a specific contribution to the three-dimensional form of the cell and to its mobility.

The driving force of the cytoskeleton is a mechanism that biologists have called dynamic instability. The cytoskeletal filaments—especially microtubules and microfilaments—are in a state of continuous flux where monomers are added to one end and taken away at the other, and the filament is growing or shortening according to which end is having the fastest run. But what is most surprising is that all this requires lots of energy, which means that the cell is investing enormous amounts of energy not in building structures but in making them unstable!

In order to understand the logic of dynamic instability, we need to keep in mind that cytoskeletal filaments are unstable only when their ends are not attached to special molecules that have the ability to anchor them. Every microtubule, for example, starts from an organizing center (the centrosome), and the extremity which is attached to this structure is perfectly stable, whereas the other extremity can grow longer or shorter, and becomes stable only when it encounters an anchoring molecule in the cytoplasm. If such an anchor is not found, the whole microtubule is rapidly dismantled and another is launched in another direction, thus allowing the cytoskeleton to explore all cytoplasm's space in a short time.

Dynamic instability, in other words, is a mechanism that allows the cytoskeleton to build structures with an exploratory strategy, and the power of this strategy can be evaluated by considering how many different forms it can give rise to. The answer is astonishing: the number of different structures that cytoskeletons can create is potentially unlimited. It is the anchoring molecules (called accessory proteins) that ultimately determine the three-dimensional forms of the cells and the movements that they can perform, and there could be endless varieties of anchoring molecules. A proof of this outstanding versatility is the fact that the cytoskeleton was invented by unicellular eukaryotes but was later exploited by metazoa to build completely new structures such as the axons of neurons, the myofibrils of muscles, the mobile mouths of macrophages, the tentacles of killer lymphocytes, and countless other specializations.

Dynamic instability, in conclusion, is a means of creating an unlimited number of cell types with only one common structure and with the choice of a few anchoring molecules. But this is possible only because there is no necessary relationship between the components of the cytoskeleton and the cellular structures that the cytoskeleton is working on. The anchoring molecules are true adaptors that perform two independent recognition processes: microtubules on one side and cellular structures on the other side. This means that they are based on arbitrary rules, on true natural conventions that have been referred to as the cytoskeleton codes (Barbieri 2003).

The Tubulin Code

Tubulin is the major component of the microtubules, the filaments that form an internal scaffolding in all eukaryotic cells and give origin to organelles such as cilia, centrioles, basal bodies, and the mitotic spindle. Most microtubules are in a state of rapid turnover and alternate very quickly between growth and shrinkage. Within the cell, however, there is also a population of microtubules that are relatively stable, in the sense that their turnover is measured in hours rather than minutes. What is particularly important is that the microtubules undergo a variety of posttranslational modifications (PTMs) that have been strongly conserved in evolution and are found in all eukaryotic taxa.

These PTMs consist in processes like acetylation, phosphorylation, polyglutamylation, polyglycylation, detyrosination, and palmitoylation that act preferentially on stable microtubules. They have been studied with various tests on purified tubulin, but the experiments have failed to detect any direct effect of the PTMs on the dynamics of the microtubules (Maruta et al. 1986; Webster et al. 1990). This means that PTMs do not act by changing directly the intrinsic properties of the microtubules, but rather by providing combinatorial signals for the recruitment of proteins that interact with the microtubules. Different combinations of PTMs, in other words, act like signposts that specify the properties that microtubules are going to have in different regions of the cell or in different periods of the cell cycle. To this set of signposts, Verhey and Gaertig (2007) have given the name of tubulin code.

An organic code, as we have seen, requires molecules that act like adaptors between two different molecular worlds. Verhey and Gaertig have identified three major classes of microtubule binding proteins that can be considered adaptors of the tubulin code: (1) microtubule associated proteins (MAPs) such as Tau, MAP1, and MAP2 that bind statically along the length of microtubules; (2) plus-end tracking proteins (+TIPs) that bind in a transient manner to the plus-ends of growing microtubules; and (3) molecular motors that use the energy of ATP hydrolysis to carry cargoes along microtubule tracks.

Verhey and Gaertig (2007, p. 2155) have also called attention to a unique characteristic of the tubulin code. Many epigenetic modifications are transmitted from one generation to the next, but this does not usually happen in the tubulin world:

> Some microtubule-based organelles (e.g., centrosomes and basal bodies) are inherited by a template-driven mechanism but there is no evidence that the template organelle directly influences the PTM pattern in the new organelle. Rather, the PTM pattern is recreated in the newly formed organelle in a gradual manner …. other microtubule-based structures, such as cytoplasmic microtubules, the mitotic spindle and cilia, are formed de novo mostly, if not entirely, from unmodified tubulin heterodimers. Thus, in case of both template-dependent and template-independent microtubular structures, PTM patterns are probably recreated without a direct influence of preexisting PTMs.

The existence of the tubulin code, in conclusion, is based on sound experimental data but the actual deciphering of its rules is still at an early stage. Luckily, the evidence in favor of this code is steadily growing and various reports have recently contributed to extending this new field of research (Janke 2014; Raunser and Gatsoiannis 2015; Barisic and Maiato 2016; Chakraborti et al. 2016).

The Compartment Codes

The cell membrane of bacteria is like a molecular skin because it synthesizes its molecules in situ, just as a skin layer contains the cells that continually renew it. In eukaryotes, instead, the cell membrane is produced by a completely different mechanism. The membrane replacements are made in the interior of the cell in the form of vesicles that move towards the surface and become incorporated into the existing membrane, while other vesicles detach themselves from the plasma membrane and move towards the interior to be recycled. In eukaryotes, in other words, the plasma membrane is the result of two opposite flows of vesicles, and its integrity is due to the perpetual motion of these ascending and descending currents.

This mechanism may appear unnecessarily complex, especially when it is compared with the simplicity of the bacterial one, but this is only a first impression. Its true logic comes immediately to light as soon as we regard it not as an isolated case, but as an example of a wider class of phenomena, more precisely, as one of the various mechanisms that eukaryotic cells employ to build their compartments. The vesicles that are destined to the plasma membrane, in fact, are produced in the golgi apparatus together with vesicles that have very different destinations. Some are delivered to lysosomes and others to secretory granules.

The golgi apparatus is involved in the terminal modification of innumerable molecules that have diverse destinations, and if every molecule had to follow a specific path, the cell simply could not cope with the immensely intricate traffic that would have to be directed. The golgi apparatus, instead, delivers to their destinations an astonishing number of molecules with only two types of vesicles: one for transporting proteins outside the cell, and the other to its interior (Farquhar 1985; Pfeffer et al. 1987). This requires only two destination signals for the vesicles, however large the number of transported proteins. On top of that, the golgi apparatus produces a third type of vesicles that do not carry any destination signal, and these are the vesicles that are programmed, by default, to reach the cell membrane. As we can see, the solution is extraordinarily efficient: with a single mechanism and only two types of signals, the cell carries an enormous amount of specific products to their destinations, and also manages to continually renew its membrane.

The golgi apparatus, however, is a transit place for only a fraction of the proteins that are produced in eukaryotic cells. The synthesis of all proteins begins in the soluble part of the cytoplasm (the cytosol), and during this first step they also receive a signal that specifies their geographic destination. The piece of the amino acid chain that emerges first from the ribosome machine—the so-called peptide leader—can contain a sequence that represents an export signal to the endoplasmic reticulum (Blobel and Dobberstein 1975; Gilmore et al. 1982). If such a signal is present, the ribosome binds to the reticulum and delivers the protein into its lumen. If the peptide leader does not carry that signal, the synthesis continues on free ribosomes, and the resulting proteins are shed in the cytosol. Of these, however, only a fraction are destined to remain there, because a peptide chain can carry, in its interior, one or more signals that specify other destinations. More precisely, there are signals for protein export to the nucleus, to mitochondria, and to other cell compartments (Blobel 1980; Kelly 1985; Robinson and Austen 1987). Proteins, in conclusion, carry with them the signals of their geographic destination, and even the absence of such signals has a meaning, because it implies that the protein is destined to remain in the cytosol.

The crucial point is that there is no necessary correspondence between protein signals and geographic destinations. The export-to-the-nucleus signals, for example, could have been used for other compartments, or could have been totally different, just like the names that are given to cities, airports, and holiday resorts. This strongly suggests that the transport of proteins across the eukaryotic compartments is based on the rules of organic codes that have been referred to as compartment codes (Barbieri 2003).

Two Different Types of Biological Codes

Definitions can be a source of endless debate, especially between different fields of research. In biology, for example, information 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 Shannon (1948).

Another difference between the two disciplines has been the definition of code. The main problem in communication engineering is the reduction, and possibly the elimination, of the noise that inevitably affects the transmission of messages, and to this end a message is submitted to two anti-noise operations that are known as source coding and channel coding.

Source coding consists in reducing the length of a message by getting rid of all unnecessary features (or redundancies) in order to minimize the exposition to noise. Channel coding is instead an anti-noise operation that is achieved not by reducing but by increasing the length of a message, in which a variety of noise-detecting operations that are known as error-correcting codes (Battail 2007, 2014) are introduced. We can get an intuitive idea of channel coding by comparing it to the familiar operation that we resort to in noisy situations when we say, for example, "Alpha, Bravo, Charlie" in place of "a, b, c."

The key point is that source codes and channel codes, collectively known as transmission codes, are operations that transform a linear sequence into another linear sequence in the same alphabet world. In the case of the genetic code, instead, a linear sequence of nucleotides is transformed into a linear sequence of amino acids that belong to a different alphabet world. The result is that the genetic code and all the other organic codes require adaptors to create bridges between the two worlds, whereas no adaptors are required in the transmission 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 genomes carry 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 rests upon Trifonov's definition that "a code is any sequence pattern that can have a biological function" or "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 such transformations.

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

The sequence codes are present in both prokaryotes and eukaryotes but the differences that exist between them do not allow us to explain why the evolution of complexity has been so different in the two primary kingdoms. In the case of the organic codes, however, the situation is substantially different.

Of all the adaptor-based codes, the prokaryotes contain only the genetic code and the signal processing codes, whereas the eukaryotes have the whole lot. The key point is that the adaptor-based codes were instrumental to the evolution of the great apparatuses of the cell (the translation apparatus, the membrane system, the splicing apparatus, the golgi apparatus, the cytoskeleton system, the histone system, and the compartment system), and this does suggest that they had a profound evolutionary role.

Evolving the Eukaryotes

In the life of any organic code there is an initial phase of code exploring followed by a long phase of code conservation. The common ancestor, for example, was a code-exploring system during the evolution of the genetic code and became a code-conservation system when that evolution was completed. The descendants of the common ancestor, on the other hand, embarked on a new journey of exploration that led them to evolve the signal transduction codes. This means that the common ancestor transmitted to its immediate descendants not only the ability to conserve the genetic code, but also the potential to further explore the coding space and bring new organic codes into existence.

This potential to generate new codes, however, was not transmitted indefinitely to all descendants. After the genetic code and the signal transduction codes, Archaea and Bacteria did not evolve any other code, whereas the ancestors of the eukaryotes continued to explore the coding space and gave origin to splicing codes, histone code, cytoskeleton codes, tubulin code, compartment codes, and sequence codes.

The key point is that this experimental fact immediately suggests 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 codes into being. But how did that happen? Why did the prokaryotes lose the ability to develop new organic codes? A natural explanation 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 where the genetic code was still evolving. A direct coupling between transcription and translation required the abolition of all intermediate steps and was achieved 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 spicing.

As a second example let us consider the histone code. The ancestral DNAs were negatively charged molecules that inevitably attracted positively charged peptides, but in order to maximize the replication rate it was necessary to remove any interposition between genes and signaling 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 maintained the same complexity ever since; the cells that did not adopt a streamlining strategy conserved the potential to evolve new organic codes and gave origin to increasingly complex systems.

The very existence of the 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 to the number of coding relationships that exist between its components, i.e., to the number of its organic codes.

The Role of Endosymbiosis

In 1866, Haeckel proposed a phylogenetic tree where the first forms of life were cells without a nucleus (which he called monera), which later generated nucleated cells (protista), which in turn gave rise to multicellular organisms. In 1883, Schimper proposed that chloroplasts had once been free-living bacteria that became incorporated by a process of internalization, or endosymbiosis, into other cells, and a few decades later this hypothesis was reformulated and extended to mitochondria by Mereschowsky (1910), by Portier (1918), and by Wallin (1927).

In the 1970s and 80s, the endosymbiosis hypothesis was forcefully reproposed by Margulis (1970, 1981), and within a few years it received overwhelming support from the experimental evidence. It was found that mitochondria and chloroplasts are still carrying fragments of their ancient circular DNA, and contain bacterial ribosomes whose molecular weight is about half that of eukaryotic ribosomes. Today it is universally acknowledged that mitochondria and chloroplasts were acquired by endosymbiosis, and their RNA sequences have shown that mitochondria were derived from alfa-proteobacteria, whereas chloroplasts are the modified descendants of cyanobacteria (Yang et al. 1985; Woese 1987).

The acquisition of mitochondria had a massive evolutionary impact because it set in motion an energy revolution that profoundly changed the course of cell history (Lane and Martin 2010; Lane 2011).

The genome of a bacterial cell gets all the energy it needs for its expression from a single bacterium, whereas the genome of a eukaryotic cell gets its energy from hundreds or thousands of symbiotic bacteria. This means that endosymbiosis allowed eukaryotes to have much more energy at their disposal, and it has in fact been calculated that "… the average eukaryote has 1,200 times as much energy per gene as the average prokaryote" (Lane 2015, p. 173). Cells spend as much as 80 % of their total energy budget in protein synthesis, and the higher the number of genes the higher the cost of protein synthesis. This is why "… there are about 13,000 ribosomes in an average bacterium such as E. coli, and at least 13 million ribosomes in a single liver cell" (Lane 2015, p. 172).

The acquisition of mitochondria, in other words, allowed the cells to enormously expand their genomes and their power. Augmenting the power of a car, on the other hand, produces a bigger and a faster car but does not change its nature. According to an influential school of thought, however, endosymbiosis did just that: it transformed the union of two prokaryotes into the first eukaryote.

This conclusion is based on two considerations: (1) the hypothesis that only Archaea and Bacteria descended from the common ancestor, and (2) the evidence that Archaea and Bacteria have not substantially changed in billions of years. The first point implies that it was Archaea and Bacteria that gave origin to eukaryotes, while the second point implies that none of them could do it individually, and only a fusion, or merger, of their cells could produce a cell that had a completely different nature.

This theory has been expressed in various ways. According to Martin and Müller, the original cell fusion took place between a methanogenic archaeon and the alfa-proteobacterium whose descendants evolved into mitochondria (Martin and Müller 1998). According to Lòpez-Garcia and Moreira (1999) the eukaryotic cell emerged from the association of a methanogenic archaeon with a myxobacterium that supplied molecular hydrogen, and subsequently this system engulfed a third partner, the endosymbiont that gave origin to mitochondria.

In both cases, the basic assumption is that the eukaryotic cell is a chimera that originated from the fusion of an archaeon with a bacterium in an extraordinary episode of transmutation that Franklin Harold described as "effectively a miracle" (Harold 2014, p. 124).

Two or Three Kingdoms?

The hypothesis that Archaea and Bacteria were the sole descendants of the common ancestor leads inevitably to the conclusion that the eukaryotes are archaeo-bacterial chimeras (the bacterial theory of life), but it is important to keep in mind that an alternative does exist: it is the idea proposed by Woese that the common ancestor gave origin to three primary kingdoms, not just two, and that the ancestors of the eukaryotes were neither Archaea nor Bacteria but a third type of cells that he called urkaryotes. It is also important to keep in mind that this proposal is compatible with the results of whole-genome phylogenies (Harold 2014) and represents therefore a legitimate scientific theory.

Despite this fact, the idea of the three kingdoms is often described as a case of eukaryotic chauvinism. According to Lane, for example, "some scientists like to view the eukaryotes as descending from the very base of the tree of life, for what I see as basically emotional reasons…the problem seems to be a case of anthropocentric dignity. We are eukaryotes and it offends our dignity to see ourselves as Johnny-come-lately genetic mongrels" (Lane 2015, p. 162).

In reality, the three kingdoms were proposed by Woese exclusively on the basis of experimental data and have been defended ever since with strictly scientific arguments.

In the case of the eukaryotic signature genes, for example, it is legitimate to say that they descended from prokaryotes and that their original prokaryotic features became transformed beyond recognition, but it is equally legitimate, and far more parsimonious, to say that they do not have prokaryotic features because they did not descend from prokaryotes.

An argument that is often raised against the three-kingdoms hypothesis is that this idea seems to imply that eukaryotic complexity arrived very early in the history of life, and that goes against the principle that in evolution simplicity came first and complexity later. But let us take a closer look at the properties can we can attribute to the urkaryotes. The complexity of the genetic code was already present in the common ancestor and was transmitted to its immediate descendants together with the potential to evolve the signal processing systems that we find in Archaea, Bacteria, and Eukarya. In this framework, the urkaryotes represent the lineage that evolved the third signal-processing system, and for this purpose they did not have to be more complex than the cells that evolved the signal processing systems of Archaea and Bacteria.

The argument from complexity, in other words, is not valid, because all the immediate descendants of the common ancestor had comparable levels of complexity. The problem is what happened to the descendants of the first cells, because in this case we know that some of them became more complex and others did not.

It is legitimate to say that the potential to evolve new organic codes was lost for good in Archaea and Bacteria, and then it was restored again by an extraordinary fusion of an archaeon with a bacterium. But it is also legitimate, and far more parsimonious, to say that that potential was not lost in the urkaryotes, and from these it was transmitted to all members of the third kingdom that eventually evolved into eukaryotes.

Life and Codes

The evolution of life took place exclusively in single cells for about 3 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, in particular with the codes that have been discovered in embryonic development and in nervous systems (Barbieri 2015b). Among them: the Hox code (Hunt et al. 1991; Kessel and Gruss 1991), the adhesive code (Redies and Takeichi 1996; Shapiro and Colman 1999), the transcriptional code (Jessell 2000; Marquard and Pfaff 2001; Altaba et al. 2003; Flames et al. 2007), the neural code (Nicolelis and Ribeiro 2006; Nicolelis 2011), a neural code for taste (Di Lorenzo 2000; Hallock and Di Lorenzo 2006), an odorant receptor code (Dudai 1999; Ray et al. 2006), a space code in the hippocampus (O'Keefe and Burgess 1996, 2005; Hafting et al. 2005; Brandon and Hasselmo 2009; Papoutsi et al. 2009), the synaptic code for cell-to-cell communication (Hart et al. 1995; Szabo and Soltesz 2015), the apoptosis code (Basañez and Hardwick 2008; Füllgrabe et al. 2010), the bioelectric code (Tseng and Levin 2013; Levin 2014), the acoustic codes (Farina and Pieretti 2014), and the glycomic code (Buckeridge and De Souza 2014; Tavares and Buckeridge 2015).

Even in these cases, however, the discovery of new codes circulated only in small circles and did not modify the traditional view of biology. As a result, we still have a theoretical framework that contemplates only two codes in nature: the genetic code that appeared at the origin of life and the codes of culture that arrived almost 4 billion years later. That amounts to saying that there have been no other codes in between, and therefore that codes are extraordinary exceptions, not normal components of life.

In reality, organic codes have appeared throughout the whole history of life and have played a crucial role in evolution. Let us take a look, for example, at the evolution of protein synthesis. The picture that emerges from the experimental data is that many apparatuses of protein synthesis evolved in the history of life, because their components are highly variable in different lineages; the ribosomes, for example, have average molecular weights of two million in prokaryotes and four million 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.

In fact, 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 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 fundamental components of life that give us a new understanding of macroevolution, and their study is destined to become an increasingly relevant part of the life sciences.

Acknowledgments I am grateful to Salthe and to two anonymous reviewers whose suggestions greatly improved the first version of this manuscript.

References

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

Adl SM, Simpson ABG, Farmer MA et al (2005) The new higher-level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 52:399–451

Agalioti T, Chen G, Thanos D (2002) Deciphering the transcriptional histone acetylation code for a human gene. Cell 111(3):381–392

Alberts B, Johnson A, Lewis J et al (2007) Molecular biology of the cell, 5th edn. Garland, New York

Altaba AR, 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

Baldauf SL (2003) The deep roots of eukaryotes. Science 300:1703–1706

Barash Y, Calarco JA, Gao W et al (2010) Deciphering the splicing code. Nature 465:53–59

Barbieri M (2003) The organic codes: an introduction to semantic biology. Cambridge University Press, Cambridge

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

Barbieri M (2015b) 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

Barghoorn ES, Tyler SM (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 JM (2008) Unravelling the Bcl-2 apoptosis code with a simple model system. PLoS Biol 6(6):e154. doi:10.137/journal.pbio.0060154

Battail G (2007) Information theory and error-correcting codes in genetics and biological evolution. In: Barbieri M (ed) Introduction to biosemiotics. Springer, Dordrecht, pp 299–345

Battail G (2014) Information and life. Springer, Dordrecht

Berger SL (2007) The complex language of chromatin regulation during transcription. Nature 447:407–412

Berridge M (1985) The molecular basis of communication within the cell. Sci Am 253:142–152

Bilwes AM, Alex LA, Crane BR, Simon MI (1999) Structure of CheA, a signal-transducing histidine kinase. Cell 96:131–141

Blobel G (1980) Intracellular membrane topogenesis. Proc Natl Acad Sci USA 77:1496–1500

Blobel G, Dobberstein B (1975) Transfer of proteins across membranes. J Cell Biol 67:852–862

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

Brandon MP, Hasselmo ME (2009) Sources of the spatial code within the hippocampus. Biol Rep 1:3–7

Brown JR, Doolittle WF (1997) Archaea and the prokaryote–eukaryote transition. Microbiol Rev 61:456–502

Buckeridge MS, De Souza AP (2014) Breaking the "glycomic code" of cell wall polysaccharides may improve second-generation bioenergy production from biomass. Bioenerg Res 7:1065–1073. doi:10.1007/s12155-014-9460-6

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

Budisa N (2014) Xenobiology, new-to-nature synthetic cells and genetic firewall. Curr Org Chem 18:936–943

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

Chakraborti S, Natarajan K, Curiel J et al (2016) The emerging role of the tubulin code: from the tubulin molecule to neuronal function and disease. Cytoskeleton. doi:10.1002/cm.21290

Cooper TA, Wan L, Dreyfuss G (2009) RNA and disease. Cell 136:777–793

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

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

Dhir A, Buratti E, van Santen MA et al (2010) The intronic splicing code: multiple factors involved in ATM pseudoexon definition. EMBO J 29:749–760

Di Lorenzo PM (2000) The neural code for taste in the brain stem: response profiles. Physiol Behav 69:87–96

Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284:2124–2129

Doolittle WF, Bapteste E (2007) Pattern pluralism and the tree of life hypothesis. Proc Natl Acad Sci USA 104:2043–2049

Dudai Y (1999) The smell of representations. Neuron 23:633–635

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

Farquhar MG (1985) Progress in unravelling pathways of golgi traffic. Ann Rev Cell Biol 1:447–488

Fitch WM, 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 DM et al (2007) Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. J Neurosci 27(36):9682–9695

Fu XD (2004) Towards a splicing code. Cell 119:736–738

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

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, Weinheim

Gamow G (1954) Possible relation between deoxyribonucleic acid and protein structures. Nature 173:318

Gilmore R, Blobel G, Walter P (1982) Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. J Cell Biol 95:463–469

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

Görlich D, Artmann S, Dittrich P (2011) Cells as semantic systems. Biochem Biophys Acta 1810(10):914–923

Görlich D, Dittrich P (2013) Molecular codes in biological and chemical reaction networks. PLoS ONE 8(1):e54694. doi:10.1371/journal.pone.0054694

Gräff J, Mansuy IM (2008) Epigenetic codes in cognition and behavior. Behav Brain Res 192:70–87

Haeckel E (1866) Generalle Morphologie der Organismen. Georg Reimer, Berlin

Hafting T, Fyhn M, Molden S et al (2005) Microstructure of a spatial map in the entorhinal cortex. Nature 436:801–806

Hallock RM, Di Lorenzo PM (2006) Temporal coding in the gustatory system. Neurosci Behav Rev 30:1145–1160

Hansen JC, Tse C, Wolfe AP (1998) Structure and function of the core histone N-termini: more than meets the eye. Biochemistry 37:17637–17641

Harold FM (2014) In search of cell history: the evolution of life's building blocks. The University of Chicago Press, Chicago

Hart AC, Sims S, Kaplan JM (1995) Synaptic code for sensory modalities revealed by *C. elegans* GLR-1 glutamate receptor. Nature 378:82–85

Hartman MCT, Josephson K, Lin C-W, Szostak JW (2007) An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. PLoS One 2(10):e972. doi:10.1371/journal.pone.0000972

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 et al (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

Jenuwein T, Allis CD (2001) Translating the histone code. Science 293:1074–1080

Jessell TM (2000) Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nat Genet 1:20–29

Jun S-R, Sims GE, Wu GA, Kim SH (2010) Whole genome phylogeny of prokaryotes by feature frequency profiles: an alignment-free method with optimal resolution. Proc Natl Acad Sci USA 107:133–138

Keeling PJ, Burger G, Durnford DG et al (2005) The tree of the eukaryotes. Trends Ecol Evol 20:670–676

Kelly RB (1985) Pathways of protein secretion in eukaryotes. Science 230:25–31

Kessel M, Gruss P (1991) Homeotic transformation of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. Cell 67:89–104

Kim J, Daniel J, Espejo A et al (2006) Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep 7(4):397–403

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 AH (2003) Life on a young planet. The first three billion years of evolution on Earth. Princeton University Press, Princeton
Komander D, Rape M (2012) The ubiquitin code. Annu Rev Biochem 81:203–229
Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. Nat Rev Microbiol 1:127–136
Koonin EV (2007) The biological big bang model for the major transitions in evolution. Biol Direct 1:22
Koonin EV (2012) The logic of chance. The nature and origin of biological evolution. Pearson Education, Upper Saddle River
Kornberg RD, Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell 98:285–294
Kühn S, Hofmeyr J-HS (2014) Is the "histone code" an organic code? Biosemiotics 7(2):203–222
Kurland CG, Collins LJ, Penny D (2006) Genomics and the irreducible nature of eukaryotic cells. Science 312:1011–1014
Lane N (2011) Energetics and genetics across the prokaryote–eukaryote divide. Biol Direct 6:35
Lane N (2015) The vital question. Energy, evolution and the origins of complex life. Norton, New York
Lane N, Martin W (2010) The energetic of genome complexity. Nature 467:929–934
Levin M (2014) Endogenous bioelectrical networks store non-genetic patterning information during development and regeneration. J Physiol 592(11):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
Lòpez-Garcia P, Moreira D (1999) Metabolic symbiosis at the origin of eukaryotes. Trends Biochem Sci 24:88–93
Maraldi NM (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
Margulis L (1970) Origin of eucaryotic cells. Yale University Press, New Haven
Margulis L (1981) Symbiosis in cell evolution. Freeman, San Francisco
Marijuán PC, Navarro J, del Moral R (2015) How the living is in the world: an inquiry into the informational choreographies of life. Prog Biophys Mol Biol 119:469–480
Marquard T, Pfaff SL (2001) Cracking the transcriptional code for cell specification in the neural tube. Cell 106:651–654
Martin W, Müller M (1998) The hydrogen hypothesis for the first eukaryote. Nature 342:37–41
Maruta H, Greer K, Rosenbaum JL (1986) The acetylation of alpha-tubulin and its relationship to the assembly and disassembly of microtubules. J Cell Biol 103:571–579
Matlin A, Clark F, Smith C (2005) Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol 6:386–398
Maurer-Stroh S, Dickens NJ, Hughes-Davies L et al (2003) The tudor domain "royal family": tudor, plant agenet, chromo, PWWP and MBT domains. Trends Biochem Sci 28(2):69–74
Mereschowsky C (1910) Theorie der Zwei Pflanzenarten als Grundlage der Symbiogenesis, einer neuen Lehre der Entstehung der Organismen. Biologisches Zentralblatt 30:278–303, 321–347, 353–367
Miller RV (1998) Bacterial gene-swapping in nature. Sci Am 278:67–71
Mujtaba S, Zeng L, Zhou MM (2007) Structure and acetyl-lysine recognition of the bromodomain. Oncogene 26(37):5521–5527
Nicolelis M (2011) Beyond boundaries: the new neuroscience of connecting brains with machines and how it will change our lives. Times Books, New York
Nicolelis M, Ribeiro S (2006) Seeking the neural code. Sci Am 295:70–77
Nudler E, Mironov AS (2004) The riboswitch control of bacterial metabolism. Trends Biochem Sci 29(1):11–17
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
O'Keefe J, Burgess N (1996) Geometric determinants of the place fields of hippocampal neurons. Nature 381:425–428
Osawa S (1995) Evolution of the genetic code. Oxford University Press, New York
Owen DJ, Ornaghi P, Yang J-C et al (2000) The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase Gcn5p. EMBO J 19(22):6141–6149
Papoutsi M, de Zwart JA, Jansma JM et al (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 SM, Salzberg SL (2007) A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. BMC Bioinform 8:159
Peterson CL, Laniel M-A (2004) Histones and histone modifications. Curr Biol 14:546–551
Pfeffer SR, Rothman JE (1987) Biosynthetic protein transport and sorting by the endoplasmic reticulum and golgi. Annu Rev Biochem 56:829–852
Portier P (1918) Les symbiotes. Masson et Cie, New York
Raunser S, Gatsoiannis C (2015) Deciphering the tubulin code. Cell 161(5):960–961
Ray A, Naters WVDG, Shiraiwa T, Carlson JR (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
Rivera MC, Lake JA (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. Nature 431:152–155
Robinson A, Austen B (1987) The role of topogenic sequences in the movement of proteins through membranes. Biochem J 246:249–261
Salgado H, Moreno-Hagelsieb G, Smith TF, Collado-Vides J (2000) Operons in *Escherichia coli*: genomic analyses and predictions. Proc Natl Acad Sci USA 97(12):6652–6657
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
Schimmel P, Giegé R, Moras D, Yokoyama S (1993) An operational RNA code for amino acids and possible relationship to genetic code. Proc Natl Acad Sci USA 90:8763–8768
Schimper AFW (1883) Uber die Entwickelung der Chlorophyllkörner und Farbkorper. Bot Ztg 41:105–114
Schopf JW (1999) Cradle of life. The discovery of Earth's earliest fossils. Princeton University Press, Princeton
Schreiber SL, Bernstein BE (2002) Signaling network model of chromatin. Cell 111:771–778
Shannon CE (1948) A mathematical theory of communication. Bell Labs Tech J 27(3):379–423
Shapiro L, Colman DR (1999) The diversity of cadherins and implications for a synaptic adhesive code in the CNS. Neuron 23:427–430
Simonson AB, Servin JA, Skophammer RG et al (2005) Decoding the genomic tree of life. Proc Natl Acad Sci USA 102:6608–6613
Snel B, Huynen MA, Dulith BA (2005) Genome trees and the nature of genome evolution. Ann Rev Microbiol 59:191–209
Solis AS, Shariat N, Patton JG (2008) Splicing fidelity, enhancers, and disease. Front Biosci 13:1926–1942
Stergachis AB, Haugen E, Shafer A et al (2013) Exonic transcription factor binding directs codon choice and affects protein evolution. Science 342:1367–1372
Strahl BD, Allis D (2000) The language of covalent histone modifications. Nature 403:41–45

Szabo GG, Soltesz I (2015) Pass-through code of synaptic integration. Neuron 87:1124–1126
Tavares EQP, Buckeridge MS (2015) Do plant cells have a code? Plant Sci 241:286–294
Tazi J, Bakkour N, Stamm S (2009) Alternative splicing and disease. Biochim Biophys Acta 1792:14–26
Trifonov EN (1989) The multiple codes of nucleotide sequences. Bull Math Biol 51:417–432
Trifonov EN (1996) Interfering contexts of regulatory sequence elements. Cabios 12:423–429
Trifonov EN (1999) Elucidating sequence codes: three codes for evolution. Ann N Y Acad Sci 870:330–338
Tseng AS, Levin M (2013) Cracking the bioelectric code. Probing endogenous ionic controls of pattern formation. Commun Integr Biol 6(1):1–8
Tucker BJ, Breaker RR (2005) Riboswitches as versatile gene control elements. Curr Opin Struct Biol 15(3):342–348
Turner BM (2000) Histone acetylation and an epigenetic code. BioEssays 22:836–845
Turner BM (2002) Cellular memory and the histone code. Cell 111:285–291
Turner BM (2007) Defining an epigenetic code. Nat Cell Biol 9:2–6
Verhey KJ, Gaertig J (2007) The tubulin vode. Cell Cycle 6(17):2152–2160
Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? Trends Genet 20(1):44–50
Wallin JE (1927) Symbionticism and the origin of species. Williams and Wilkins, Baltimore
Wang Z, Burge C (2008) Splicing regulation: from a part list of regulatory elements to an integrated splicing code. RNA 14:802–813
Wang GS, Cooper TA (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet 8:749–761
Weatheritt RJ, Babu MM (2013) The hidden codes that shape protein evolution. Science 342:1325–1326
Webster DR, Wehland J, Weber K, Borisy GG (1990) Detyrosination of alpha tubulin does not stabilize microtubules in vivo. J Cell Biol 111:113–122
Woese CR (1965) Order in the genetic code. Proc Natl Acad Sci USA 54:71–75
Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271
Woese CR (2000) Interpreting the universal phylogenetic tree. Proc Natl Acad Sci USA 97:8392–8396
Woese CR, Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 74:5088–5090
Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eukarya. Proc Natl Acad Sci USA 87:4576–4579
Wolfe AP, Hayes JJ (1999) Chromatin disruption and modification. Nucleic Acid Res 27:711–720
Wu J, Grunstein M (2000) 25 years after the nucleosome model: chromatin modifications. Trends Biochem Sci 25(12):619–623
Yang D, Oyaizu Y, Olsen GJ, Woese CR (1985) Mitochondrial origins. Proc Natl Acad Sci USA 82:443–447