Semiotic modelling of biological processes: genetic information system

João Queiroz[a,b,c] & Charbel El-Hani[b,c,d]

a.Institute of Arts and Design, Federal University of Juiz de Fora, Brazil.
b.Research Group in History, Philosophy, and Biology Teaching, Institute of Biology, Federal University of Bahia, Brazil.
c.Graduate Studies Program in History, Philosophy, and Science Teaching, Federal University of Bahia/State University of Feira de Santana, Brazil.
d.Graduate Studies Program in Ecology and Biomonitoring, Federal University of Bahia, Brazil.

General abstract: Here we introduce biosemiotics as a field of research that develops models of life processes focusing on their informational aspects. Peirce's general concept of semiosis can be used to analyze such processes, and provide a powerful basis for understanding the emergence of meaning in living systems, by contributing to the construction of a theory of biological information. Peirce's theory of sign action is introduced, and the relation between 'information processing' and sign processes is discussed, and, in fact, a semiotic definition of information is proposed. A biosemiotic model of genetic information processing in protein synthesis is developed.

As we have argued in previous works (El-Hani, Queiroz & Emmeche 2009, 2006; Queiroz & El-Hani 2006; Queiroz, Emmeche & El-Hani 2005), the use of Peircean semiotic concepts to interpret 'information talk' in biology can significantly contribute to the construction of a coherent account of meaning and information in living systems. This is an important task, since biology is pervaded by informational concepts, particularly in fields with important social and technological implications, such as genetics, molecular biology, and genomics, but currently lacks a theory of biological information that can sufficiently account for its semantic and pragmatics aspects (Griffiths, 2001; Jablonka, 2002; El-Hani, Queiroz & Emmeche, 2009).

In this lecture, we introduce the Peircean notion of information and then applied this model to studies about the genetic information system.

1. Meaning, information and semiosis

The notions of 'meaning', 'information', and 'semiosis' intersect in different ways (Johansen 1993). Debrock (1996) comments that Peirce defined 'information' at least ordinarily (CP 2.418), metaphysically (CP 2.418), as a connection between form and matter, and logically (W 1.276), as the product of extension and intension of a concept. We previously argued that definitions of Sign found in Peirce lead to a definition of information as the communication of a form from O to I through S (Queiroz, Emmeche & El-Hani 2005; El-Hani, Emmeche & Queiroz 2006). In these terms, it amounts to the communication of a habit embodied in the Object to the Interpretant, so as to constrain (in general) the Interpretant as a Sign or (in semiotic systems) the interpreter's behavior. To put it differently, the production of an effect of the Sign on the interpreter results from the communication of the form embodied in the Object (as a regularity), via Sign, to the Interpretant.

According to this approach, ‘information’ can be strongly associated with the concepts of ‘meaning’ and ‘semiosis’.

It is important to emphasize that the form communicated from the Object to the Interpretant through the Sign is a regularity, a habit that allows a given semiotic system to interpret that form as indicative of a class of entities, processes, phenomena, and, thus, to answer to it in a regular way. Otherwise, the semiotic system would not be really capable of interpreting the Object by means of its effect on it (Interpretant), mediated by a Sign.

Peirce’s (CP 8.177) idea that a Sign determines an Interpretant in some ‘actual’ or ‘potential’ Mind (in other passages, a “quasi-mind”; see CP 4.536) also plays an important role in our arguments. On the grounds of this idea, we differentiate between ‘potential’ and ‘effective’ semiosis. We understand potential semiosis as a triadically-structured process that might be, but it is not effectively taking place at a given time t. Effective semiosis, in turn, is a Sign in effective action, i.e., a Sign that, by being actualized, has an actual effect on the interpreter. Following the distinction between potential and effective semiosis, we can define potential and effective information (see below).

According to our interpretation of Peirce’s ideas, information has a processual nature: it is a process of communicating a form to the Interpretant that operates as a constraining influence on possible patterns of interpretative behavior. When applying this general semiotic approach to semiotic systems, information will most often be an interpreter-dependent objective process. It cannot be dissociated from the notion of a situated agent. It is interpreter-dependent in the sense that information triadically connects representation (Sign), Object, and an effect (Interpretant) on the interpreter (which can be an organism or a part of an organism). The form – as a regularity embodied in the Object – acts as constraint on the interpreter’s behavior. In sum, information in a semiotic system depends on both the interpreter and the Object (in which the form communicated in information is embodied as a constraining factor of the interpretative process).

A framework for thinking about information as a process can be built in Peircean terms by employing the following definitions:

> [Information = semiosis] A triadic-dependent process through which a form embodied in the Object in a regular way is communicated to an Interpretant through the mediation of a Sign.
>
> [Potential information = potential semiosis] A process of communicating a form from an Object to an Interpretant through the mediation of a Sign that could take place in a given moment.
>
> [Effective information = effective semiosis] The process by which a Sign effectively produces an effect (Interpretant) on some semiotic system (an interpreter) by making the Interpretant stand in a similar relation to something else (the Object of the Sign) as that in which the Sign itself stand.

Thus, the Sign mediates the relation between Object and Interpretant. The Sign effectively communicates, in this way, a form from the Object to the Interpretant, changing the state of the interpreter.

2. Information talk in biology

During the 1950s and 1960s, genetics, cytology, and molecular biology have been swamped by terms borrowed from information theory. This 'information talk', or 'quasi-semiotics', still pervades these fields, including widely used terms such as 'genetic code', 'messenger RNA', 'transcription', 'translation', 'transduction', 'recognition', 'genetic information', 'chemical signals', 'cell signaling' etc. But, as the concept of information and its plethora of associated notions were introduced in biology, so were several problems with which the tradition of biology was unprepared to cope. Instead of deepening the discussion about the problems involved in information talk, the trend in the biological sciences has been one of treating 'information' as merely sequence information in DNA or proteins.

A number of researchers consider information talk as inadequate and 'just metaphorical', thus expressing a skepticism about the use of the term 'information' and its derivatives in biology (Stuart 1985, Sarkar 1996). We disagree with this position, claiming instead that the notion of information and other related ideas grasp some fundamental features of biological systems and processes that might be otherwise neglected. The concepts of 'code', 'information', 'signals', 'message', 'signaling' and so on can be seen as necessary to understand the organization of relations in living beings in such a way that makes it clear that what happens in such beings is much more than simple chemistry. Bray, for instance, argued that "organisms can be viewed as complex information-processing systems, where molecular analysis alone may not be sufficient" (cited by Williams 1997, p. 476-477). Ideker, Galitski and Hood (2001), in a paper about systems biology, argue that biology is an informational science. Indeed, since the early applications of cybernetic models in life sciences, biology has been increasingly conceptualized as a communication and information science (e.g., Keller 2005), even though in many cases it is not clear at all what is meant by 'information' in biology (Griffiths 2001, Jablonka 2002, Jablonka & Lamb 2005).

It is not surprising that biologists felt the need to talk about 'information' when delving more and more into the molecular micro-structure of living systems. Life scientists needed a way of conveying the idea that, even though all cellular processes are physicochemical processes, more than just physics and chemistry is going on there. They are complexly organized physicochemical processes interwoven in communication and information networks. In this context, it is quite difficult to see what would be the real advantage of stripping off information talk from biology, instead of making it more precise and exploring its consequences in more depth. Thus, the problem is not to get rid of information talk, but rather to clarify it by using a proper theoretical framework.

As Griffiths (2001) sums up, 'genetic information' is a metaphor in search of a theory. We believe this applies in general terms to information talk in biology. One possibility for building a theory of information in biology is to rely on the mathematical theory of communication. This theory allows one to define the

amount of information as the measure of the probability of selection of a particular message among the set of all possible messages. The probabilistic measure of information provided by this theory is non-semantic, indifferent to meaning (Shannon & Weaver 1949, Cover and Thomas 1999, Jablonka, 2002). It is true that this meaning-free concept of information can be useful in biological research for several purposes (Adami 2004). Nevertheless, it has been argued that such a non-semantic (and quantitative) understanding of information is not sufficient for a theory of biological information, and should be complemented by a semantic, pragmatic (and more qualitative) approach. Jablonka (2002), for instance, uses an example where a DNA sequence encoding a functional enzyme and a same-length sequence coding for a completely non-functional polypeptide (which can even have only a single different nucleotide) would contain, according to the above-mentioned measure, the same amount of information. It is obvious, however, that these two messages do not mean the same to the cell. This indicates the necessity of a treatment of information in biology that includes a semantic and a pragmatic dimension. Or, to put it differently, a theory of biological information should deal also with the meaning of 'messages' and the context in which they are interpreted. Here, we use semiotic concepts to build a semantic and pragmatic account of biological information. In particular, we propose a model of information as semiosis, grounded in Peirce's pragmatic theory of signs.

3. The gene concept and its problems

The gene concept has certainly been one of the landmarks in the history of science in the 20th century. Keller (2000), for instance, refers to the 20th century as 'the century of the gene'. Grós (1989), by his turn, claims that we live in a 'civilization of the gene'.

The term 'gene' was introduced by the Danish geneticist W. L. Johannsen, who regarded it as a kind of accounting or calculating unit, a very handy term but with no material counterpart that could be related to it with any degree of confidence (Johannsen 1909, *cf*. Falk 1986). Indeed, in the beginnings of genetics, an instrumentalist view about the status of 'gene', as a theoretical concept, prevailed (Falk 1986). The 'gene' was often regarded as nothing but a useful abstract concept to express regularities in the transmission of phenotypic traits.

Nevertheless, a realist, material view about the status of 'gene' was also found in classical Mendelian genetics. Herman J. Muller, for example, advocated the idea that genes were material units in their own rights, even though they could only be recognized through their effects. As Falk (1986) convincingly argued, the tension between instrumentalist and realist attitudes towards the status of the gene concept resulted in a fertile dialectics, described by him as a development on the pattern of 'Russian dolls', in which discoveries about the chemical nature of the gene led, in turn, to the elaboration of new functional definitions, which, in turn, led to the investigation of a deeper structural meaning, which, in turn, led to a still deeper level of functional meaning, and so on.

Genes were regarded in classical genetics as units of recombination, function, and mutation. However, as a result of the development of the understanding of the gene on the pattern of 'Russian dolls', it became eventually clear that genes were not units of either recombination or mutation. In the end, the prevailing meaning of the term in the 20th century was that of a gene as a 'unit of function'. But, after the proposal of the double helix model and the flourishing of molecular biology, the gene was redefined as a material entity, concretely existent in DNA, and it became usual to think of the gene also as a structural unit. Finally, the introduction of an informational vocabulary in molecular biology and genetics resulted in the so called 'information talk', and genes came to be often regarded also as informational units, leading to what has been called the informational conception of the gene (Stotz et al., 2004), a very popular notion in textbooks, in the media, and in public opinion. What is meant by 'information' in this case is merely sequence information in DNA or proteins (Sarkar 1998), an idea we will challenge throughout this lecture.

With the proposal of the double helix model of DNA by James Watson and Francis Crick in 1953, a realist view about the gene prevailed, DNA was established as the material basis of inheritance, and the road to the so called classical molecular gene concept was paved. Indeed, the *classical molecular gene concept*, according to which *a gene is a sequence of DNA that encodes a functional product, a polypeptide or an RNA*, can be seen as an outgrowth of the advances of molecular biology in the 1950s and 1960s. Genes seemed to be reducible, then, to concrete entities at the molecular level, namely, strings of DNA, and the structural and functional definitions of the gene were focused on a single entity (Stotz et al. 2004), resulting in a model with remarkable heuristic power.

The classical molecular gene concept is closely connected with the 'central dogma of molecular biology', conceived as a statement about the 'flow' of 'information' in a cell. In a manner that dramatically shows the strong reductionistic tendency that marked molecular biology since its beginnings (although this science seems to be gradually adopting a less reductionist view in recent years), the very idea of the dogma was that DNA makes RNA, RNA makes proteins, and proteins make the organism (see Crick 1958). But Crick also expressed the dogma more carefully as follows: 'once [sequential] information has passed into protein, it cannot get out again' (Crick 1958, [our insertion]). This 'dogma' became one of the elements in the hard core of molecular biology as a research program. In this context, the problem that no clear conception of 'information' is available in biological thought becomes quite central to molecular biology.

Since the beginnings of molecular biology, 'information' was conflated or simply identified with a substance, a string of DNA constituting a 'gene'. When information is substantialized as sequences of nucleotides in DNA, we find ourselves in a difficult position to identify other kinds of information in a cell or even in the organism as a whole. Even if we point out to other 'informational' molecules, such as RNAs and proteins, the 'information' they

allegedly ‘contain’ or ‘carry’ can be directly traced down to DNA, through the central dogma of molecular biology. When information is thus substantialized, DNA becomes a sort of reservoir from where all ‘information’ in a cell flows and to which it must be ultimately reduced. Our understanding becomes, so as to say, seduced by this purported ‘information reservoir’ and we tend, then, to overplay the role of DNA in cell systems, turning it into a complete ‘program for development’ or an all-powerful ‘controller’ of cell metabolism. But, as we are enchanted by this quite controversial picture of the role of DNA, we simply forget that DNA seems to play the role of a set of data rather than that of a program in cell systems (Atlan and Koppel 1990). Or, to put it differently, that DNA is a source of materials for cells, playing roles that are *obviously important*, but cannot be correctly described as that of a sort of master agent (or master molecule) in cell processes (Nijhout 1990). It is not DNA that *does things to* the cell; rather, it is the cell which *does things with* DNA.

The widespread usage of the informational conception of the gene makes the consequences of the substantialization of information as a set of entities in DNA go far beyond conceptual issues in genetics and molecular and cell biology. Oyama ([1985]2000) identifies a connection between this way of rendering the notion of ‘genetic information’ and genetic determinism, which has important consequences for the public understanding of science and a whole series of social, economical, and political issues related to the knowledge and applications in the fields of genetics and molecular biology.

In sum, it is an important task to clarify the concept of information in biology. But to reach any worthy result in this task, we should employ appropriate conceptual and methodological tools. Biosemiotics (for introductions to biosemiotics, see, e.g., Barbieri 2007, Hoffmeyer 1996, Kull 1999), still a somewhat neglected perspective in current debates about the gene concept, offers a theoretical ‘toolbox’ for dealing with the notion of information in biology that can help us reach a precise and coherent understanding of this central notion. We also believe biosemiotics makes it possible to formulate the notion of genetic information in a manner which does not lend support to genetic determinism.

With regard to the gene concept, several discoveries in molecular biology, including transposons, split genes, alternative splicing, consensus sequences, overlapping and nested genes, mRNA editing, transplicing, etc., posed very difficult problems to the classical molecular gene concept. These discoveries led, in Falk’s (1986:164) words, to “... an age of anarchy in the instrumental formulation of genetic entities”, in which a great number of heterodox entities was admitted into the “expanding zoo of genetic units”. It was realized that the gene is neither discrete — there are overlapping and nested genes —, nor continuous — there are introns within genes —; it does not necessarily have a constant location — there are transposons —, nor a clearcut function — there are pseudogenes; it is neither a unit of function – there are alternatively spliced genes and genes coding for multifunctional proteins –, nor a unit of structure – there are many kinds of *cis*-acting sequences affecting transcription (promoters, enhancers, terminators, etc.), split genes,

etc. – (*cf*. Falk 1986:169, Fogle 1990:356-363). In this scenario, the question 'What is a gene, after all?' became a topic of debate in the philosophy of biology (for reviews about these discoveries and the problems they bring to the gene concept, see, for instance, Falk 1986; Portin 1993; Keller 2000; Fogle 1990, 2000, El-Hani, 2007; El-Hani, Queiroz & Emmeche, 2009).

But it is not only in philosophy of biology that we find a growing recognition of the problems surrounding the gene concept. Doubts about the status of this concept are also found in empirical papers within molecular biology (possibly indicating a crisis in molecular biology as a 'normal science'). To quote just two examples, we find Wang et al. (2000), in a study of the origin of a particular gene and the complex modular structure of its parental gene, claiming that this structure "... manifests the complexity of the gene concept, which should be considered in genomic research" (*ibid.*: 1294), for instance, when one tries to predict a gene from genome data (*ibid.*: 1300). Kampa et al. (2004), by their turn, considers that their findings in an in-depth analysis of the transcriptome (the set of all transcripts of a cell) of human chromosomes 21 and 22 "… strongly support the argument for a re-evaluation of the total number of human genes and an *alternative term for* '*gene*' to encompass these growing, novel classes of RNA transcripts in the human genome" (*ibid.*: 331; emphasis added). Although they do not suggest that we should abandon the term 'gene' altogether (as, for instance, Keller 2000; see below), they comment that "... the use of the term "gene" to identify all the transcribed units in the genome may need reconsideration, given the fact that this is a term that was coined to denote a genetic concept and not necessarily a physical and measurable entity. With respect to the efforts to enumerate all functional transcribed units, it may be helpful to consider using the term 'transcript(s)' in place of gene" (Kampa et al. 2004: 341).

In the last three decades, a realist, material view of the gene has been superseded by a pluralist view that was captured by Falk in the following statement: "Today the gene is not *the* material unit or *the* instrumental unit of inheritance, but rather *a* unit, *a* segment that corresponds to *a* unit-function as defined by the individual experimentalist's needs" (Falk 1986: 169. Emphasis in the original). Ambiguities have been a feature of the gene concept throughout its whole history (Kitcher 1982, Falk 1986) and they even have been heuristically useful in the past. Even though Falk is uncertain as to whether or not the current ambiguities will also be helpful, he does not seem to consider the sort of attitude he describes in current scientists as a reason to lose our hope regarding the status of the gene concept (see Falk 1986, 2000, 2001). Other researchers, however, consider that the conceptual variation currently observed in the case of the gene can lead to confusion (e.g., Fogle 1990, 2000).

As 20th century came to a close and we entered what seems to be a whole new era in biological research, the future of the gene didn't look bright for some thinkers. Keller (2000), for instance, considers the gene a concept "in trouble" and suggests that maybe the time is ripe to forge new words and leave that concept aside. Although some authors agreed with Keller's proposal (e.g., Rios 2004), it has not found wide acclaim; rather, it was rejected by

many reviewers of her book, such as Coyne (2000), Magurran (2000), Maynard Smith (2000b), Hall (2001), and Wilkins (2002). Recently, Keller published a paper reexamining her previous ideas under the light of new developments in molecular biology, genomics, and related areas (Keller, 2005). In this paper, she takes a more optimistic view about the future of the gene, considering her previous arguments for the need to move on to a "century beyond the gene" (ibid., p. 3) an "impassioned" one, and arguing that the 21st century will be the century of genetic systems, rather than of the gene. She does not claim anymore that the gene concept should be abandoned, but rather considers that the challenges currently posed by biological complexity demands "new ways of talking" (ibid., pp. 8-9). In order to address the interactions between the parts of living systems and the dynamics of these interactions, biologists should overcome, Keller argues, "ingrained habits of thought and speech that give ontological priority to those parts". These habits are particularly problematic in genetics, "where the parts are taken to be genes", while "genes, by definition, do not have meaning in isolation". Keller goes on to treat the cell as "a meaning making system that turns nucleotide sequences into genes". In this picture, the gene concept can survive in the 21st century, she argues, but "by reconceptualizing them as verbs" (ibid., p. 9), an idea much in agreement with the consequences of our Peircean semiotic analysis of genes.

Other philosophers of biology and also practicing scientists foresee a brighter future for the gene concept. Even though Falk admits that the gene is a concept "in tension" (Falk 2000), he seeks ways to 'save' it (Falk 2001). Waters is even more optimistic, considering that different definitions of the gene can be unified by a concept with a number of 'open' clauses, such as that of "a gene for a linear sequence in a product at some stage of genetic expression" (Waters 1994:178). Hall (2001) is also optimistic, arguing that, despite published obituaries (Gray 1992, Neumann-Held 1999, Keller 2000), the gene is not dead, but alive and well, even though 'orphaned', 'homeless', and seeking a haven from which to steer a course to its 'natural' home, the cell as a fundamental morphogenetic unit of morphological change in development and of evo-devo (the interface between evolution and development).

The attempts to save the gene also led to distinctions between different concepts, as, for instance, Griffiths and Neumann-Held's (1999) distinction between the 'molecular gene' and the 'evolutionary' gene, and Moss' (2001, 2003) distinction between gene-P (the gene as a determinant of phenotypes or phenotypic differences) and gene-D (the gene as a developmental resource). Moss forcefully argues that genes can be productively conceived in these two different ways, "albeit with nothing good resulting from the conflation of the two" (Moss 2001:85). Gene-P, on the one hand, is the "… expression of a kind of instrumental preformationism" (*ibid*:87), showing its usefulness due to the epistemic value of its predictive power and its role in some explanatory games of genetics and molecular biology. In these terms, Moss doesn't attack the much criticized construct of the 'gene for' one or another phenotypic trait, recognizing its value for some theoretical and empirical tasks. Rather, the focus of his criticism is on the tendency to

conflate this first conceptualization of the gene with a second one, that of gene-D. A gene-D is conceived, in a more realist tone, as a developmental resource defined by a specific molecular sequence and functional template capacity, which plays an entirely different explanatory role, in comparison to that of gene-P. Gene-P and gene-D are, in short, distinct concepts with different conditions of satisfaction for what it means to be a gene.

We focus on gene-D in the present lecture. Our task here is to begin the construction of a theoretical framework for a semiotic analysis of the concepts of 'gene' and 'information', on the grounds of a case study about protein-coding genes. We should emphasize the originality of this approach, not only in the specific context of molecular biology, but also in the general context of biosemiotics. We think it is important to develop biosemiotics by providing new sets of modeling tools and some exemplars or case studies to understand the precise sense in which specific life processes can be conceived as involving the action of Signs, as generally claimed by biosemioticians. Furthermore, by applying the formal notion of semiosis to model some aspects of the genetic information system, we intend to produce a radically new explanation of 'genetic information' as a semiotic process. In this effort, we will move towards a reinterpretation of what is information in a cell that hopefully avoids a number of problems detected in information talk not only in biology but also in science as a whole.

4. Some basic notions about the genetic information system

It suffices for the analysis we perform here to present some very general notions about transcription, mRNA splicing, and protein synthesis. We will deliberately avoid introducing a large number of details, which can be easily found in any molecular and cell biology textbook (*e.g.*, Griffiths et al. 1999, Lodish et al. 2003, Alberts et al. 2002, Lewin 2004).

Let us consider first a very simple model of the process of gene expression. During the synthesis of pre-mRNA, the four-base language of DNA (as a sequence of nucleotides including the bases adenine, A, guanine, G, cytosine, C, and thymine, T) is copied or 'transcribed' into the four-base language of RNA (with uracil, U, replacing T). Transcription results in functional mRNAs (messenger RNA), rRNAs (ribosomal RNA), tRNAs (transfer RNA), snRNAs (small nuclear RNA), and scRNAs (small cytoplasmic RNA), but we will focus here on the synthesis of mRNA. Other functional RNAs which play important roles in various steps in DNA processing will be mentioned in passing.

During transcription, one DNA strand acts as a 'template', determining by base pairing the order in which monomers (ribonucleoside triphosphates) are assembled to form a complementary RNA polymer, by a polymerization reaction catalyzed by the enzyme RNA polymerase.

The effects of a protein-coding gene on a given cell or organism are regulated mainly by control of gene expression at the level of transcription initiation. The transcription of a gene can be either *repressed*, when the corresponding mRNA and encoded protein or proteins are synthesized at low rates or not

synthesized at all, or *activated*, when both the mRNA and encoded protein or proteins are, *ceteris paribus*, produced at much higher rates. Through the control of gene expression, only a subset of all genes present in any cell type in a multicellular organism is really expressed. Thus, from all the potential protein products a given cell type might have, only a specific number and variety will be present. This is the fundamental basis for cell differentiation in multicellular organisms.

In the end of the 1970s, it was found that eukaryotic genes are split into pieces of coding sequence, named 'exons', separated by non-coding segments, named 'introns' (after Gilbert 1978). The discovery of split genes was one of the challenging discoveries that eventually led to the current debates about the gene concept. Now, it is well known that introns are common in multicellular eukaryotes, uncommon in many unicellular eukaryotes, and extremely rare in eubacteria and archaea. The vast majority of genes in multicellular eukaryotes contain multiple introns and the presence of such introns allows the expression of multiple related proteins from a single stretch of DNA by means of a process known as 'alternative splicing', which poses yet another challenge to the gene concept.

In eukaryotic protein-coding genes, introns are excised from a long 'primary transcript' (precursor mRNA or pre-mRNA), *i.e.*, the RNA copy of an entire DNA sequence containing both exons and introns, in a process known as RNA 'processing', which includes other events not described here. After the introns are excised, the coding exons are joined back together into a functional mRNA, which will be transported to the cytoplasm of the eukaryotic cell, where protein synthesis will take place.

Alternative splicing is rather common in mammalian genomes. Genome-wide analyses indicate that 35-59% of human genes produce alternatively spliced forms (Modrek & Lee 2002). Even though a significant portion of the predicted splicing variants are not functional, resulting from aberrant rather than regulated splicing, and, therefore, the frequencies of alternatively spliced gene products mentioned above are probably overestimated (Sorek et al. 2004), it is still the case that alternative splicing should be regarded as one of the most significant components of the functional complexity of the genome of our and many other species (Modrek & Lee 2002).

Alternative RNA splicing requires that the conceptualizations of genes move far beyond the simple scheme captured in formulas such as 'one gene-one protein or polypeptide'. One might argue, however, that such a challenge to the gene concept can be easily assimilated by simply replacing this formula by a new one, for instance, 'one gene-many proteins or polypeptides'. However, the situation is not so simple. As Keller (2000) argued, the situation is such that it does not allow us to be clear about where is the gene after all. For instance, should we call a 'gene' that piece of sequence in DNA that can generate dozens of different proteins? Or should we apply this concept to each individual spliced mRNA by formulating such an idea as that of one mature mRNA-one protein? If we opt for the second alternative, a number of other problems will follow. For instance, the mRNA molecule itself can be

further modified (RNA editing) and the final transcript can be assembled from exons derived from different pre-mRNAs (trans-splicing). More importantly, mRNAs are structures much more transient than quite basic (and, arguably, correct) intuitions about genes and their stability through generations require.

Alternative RNA splicing is an important mechanism for the production of different forms of proteins (isoforms) by different cell types. The fibronectin (FN) gene, for instance, generates more than 20 different FN isoforms. The FN gene has approximately 75,000 nucleotides (75-Kb) and contains numerous exons. After the FN pre-mRNA is transcribed from DNA, it undergoes cell type-, development- and age-specific splicing. Each FN isoform is encoded by a differently, alternatively spliced mRNA, and, therefore, each isoform results from a unique combination of exons found in the FN gene (see Figure 1).

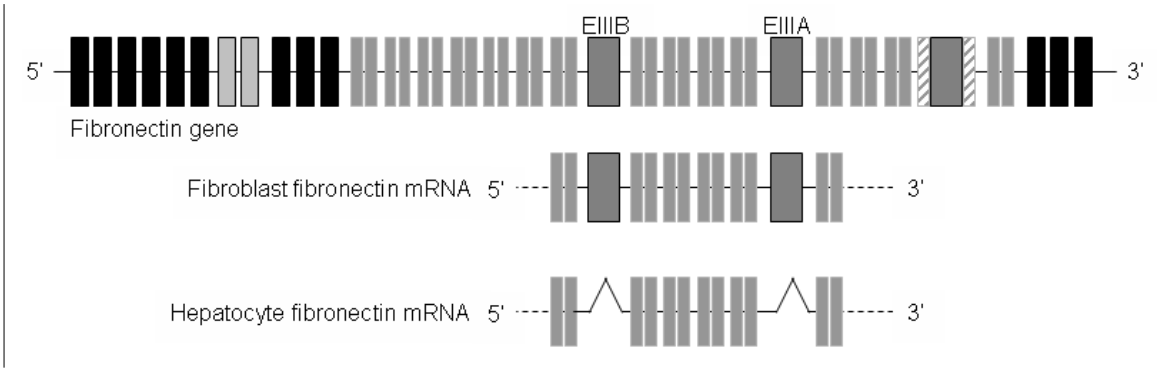

Figure 1. Cell type-specific splicing of fibronectin pre-mRNA in fibroblasts and hepatocytes. The 75-kb FN gene (top) contains multiple exons. Introns are shown in the diagram as thin lines and are not drawn to scale. Most of the introns are much longer than any of the exons. The FN mRNA produced in fibroblasts includes the EIIIA and EIIIB exons, whereas these exons are spliced out of FN mRNA in hepatocytes (modified from Lodish et al. 2000. Available at http://www.ncbi.nlm.nih.gov/PubMed/).

The combinations of exons in each isoform change its causal dispositions. This can be clearly seen in the case of the splicing of FN pre-mRNA in fibroblasts and hepatocytes. In fibroblasts, splicing of the FN pre-mRNA results in mRNAs containing exons EIIIA and EIIIB. The fibroblast FN isoform contains amino acid sequences that bind tightly to proteins in the plasma membrane, ascribing it specific causal dispositions. This specific FN isoform contributes to the adhesion of fibroblasts to the extracellular matrix. In hepatocytes, the major cell type in the liver, cell-type specific splicing results in functional FN mRNAs lacking exons EIIIA and EIIIB. As in the case of fibroblasts, we have here a FN isoform with specific causal dispositions. First, it does not show the causal dispositions the fibroblast isoform shows: FN secreted by hepatocytes does not adhere tightly to fibroblasts or most other cell types. The lack of such causal dispositions is very important to the functionality of this FN isoform, since it allows it to freely circulate in the blood stream. Nevertheless, when the wall of a vase is ruptured, hepatocyte FN plays a fundamental role in the formation of blood clots, showing its specific causal disposition, which result from the presence in the protein of fibrin-binding domains, amino acid sequences that bind to fibrin, one of the main constituents of blood clots. When hepatocyte FN is bound to fibrin, it shows yet another causal

disposition, interacting with integrins, cell-adhesion protein molecules found in the membranes of activated platelets. As a result, the blood clot is expanded through the addition of platelets.

The effects of genes on the functioning of a cell or organism can also be regulated by means of alternative pre-mRNA splicing, so as to produce different gene products from the same pre-mRNA. Particularly remarkable examples of genetic regulation at the level of RNA splicing are found, for instance, in the sex determination pathway of *Drosophila* (for a review, see, *e.g.*, Black 2003).

Finally, translation is an essential part of protein synthesis, consisting in the process by which the nucleotide sequence of an mRNA serves as a template for the synthesis of a polypeptide chain, *i.e.*, for a series of events in which amino acids are ordered and joined to form the primary structure of a protein. Three types of RNA molecules are involved in translation, performing different but cooperative functions. mRNAs are the 'vehicles' of the genetic information transcribed from DNA. The 'message' at stake is 'written' in the form of a series of three-nucleotide sequences, called 'codons', each of which specifying a particular amino acid. tRNAs play a fundamental role in the process of deciphering the codons in mRNA. Each type of amino acid has its own subset of tRNAs. They act as transporters, binding amino acids and carrying them to the growing end of a polypeptide chain in response to specific codons in the mRNA. The reason why the correct tRNA with its attached amino acid is selected at each step in protein synthesis lies in the fact that each specific tRNA molecule contains a three-nucleotide sequence, called an 'anticodon', that base-pairs with its complementary codon in the mRNA. In this manner, for each specific codon in mRNA a specific amino acid, carried by a specific tRNA, is included in a polypeptide chain, according to the rules expressed in the almost universal 'genetic code'. Along with 100 different proteins, several types of rRNA are components of ribosomes, the complex and large macromolecular structures that act, so as to say, as guides to coordinate the assembly of the amino acid chain of a protein. In fact, an rRNA (a ribozyme), and not a protein, is probably the catalyst involved in the formation of peptide bonds in protein synthesis.

Translation involves three stages: initiation, when ribosomal units assemble near the translation start site in the mRNA with the tRNA carrying the amino acid methionine base-paired with the start codon, most commonly AUG; chain elongation, in which a four-step cycle is repeated, involving the binding of a tRNA carrying an amino acid, the release of the tRNA involved in the previous step in the elongation, transfer of the growing polypeptide to the incoming amino acid catalyzed by one of the rRNAs, and translocation of the ribosome to the next codon in the mRNA; and termination, in response to stop codons UAA, UGA, and UAG.

Recognition of a codon in mRNA specifying a given amino acid by a particular tRNA is, in fact, the second step in 'decoding' the genetic 'message'. The first step is the attachment of the appropriate amino acid to a tRNA in a reaction catalyzed by a specific aminoacyl-tRNA synthetase. The specificity of the

attachment between amino acids and tRNAs results from the capacity of each one of these enzymes of recognizing *one* amino acid and *all* its compatible, or 'cognate', tRNAs. Therefore, the rules captured in the genetic code ultimately depend on the recognition activity of aminoacyl-tRNA synthetases.

Although the terms 'translation' and 'protein synthesis' are usually employed interchangeably, this is not correct, since, although translation is obviously an essential step in protein synthesis, this process involves further steps. Polypeptide chains undergo post-translational folding and often other changes, as, for instance, chemical modifications and association with other polypeptide chains, which are required for production of functional proteins. All these steps in protein synthesis can undergo regulation.

If we now check the terms presented within commas in the previous paragraphs, we will be able to see 'information talk' in action. Our strategy was to use terms that are frequently employed in the same manner in biological papers and textbooks, in order to highlight the importance of building a theory to give a precise meaning to these rather metaphorical language. As we mentioned above, Griffiths (2001) said that genetic information is a metaphor in search of a theory. An analysis of molecular biology textbooks (Pitombo et al., 2008) shows that this is really so, since no concept of 'information' is offered in those textbooks as a ground for understanding the information talk that pervades them.

5. Semiotic analysis of genes and genetic information

After this basic background for understanding the genetic information system, we can move on to an analysis of genes and genetic information grounded on Peirce's theory of signs (the original sources for this analysis are Queiroz et al. 2005, El-Hani et al., 2006, 2009. In these works, the readers will also find a more detailed analysis of transcription and translation). From the perspective of this theory, the action of a gene as a Sign should be understood as a relationship between three elements (Figure 2). By employing the definition of 'information', genetic information can be described as a semiotic process. From this perspective, there is more to genetic information than just the sequence of nucleotides in a stretch of DNA. This is an important conclusion, since it goes against the treatment of genetic information as merely sequence information in DNA or proteins, and indicates a different path to conceptualize information, in a theory of biological information grounded on Peircean semiotics.

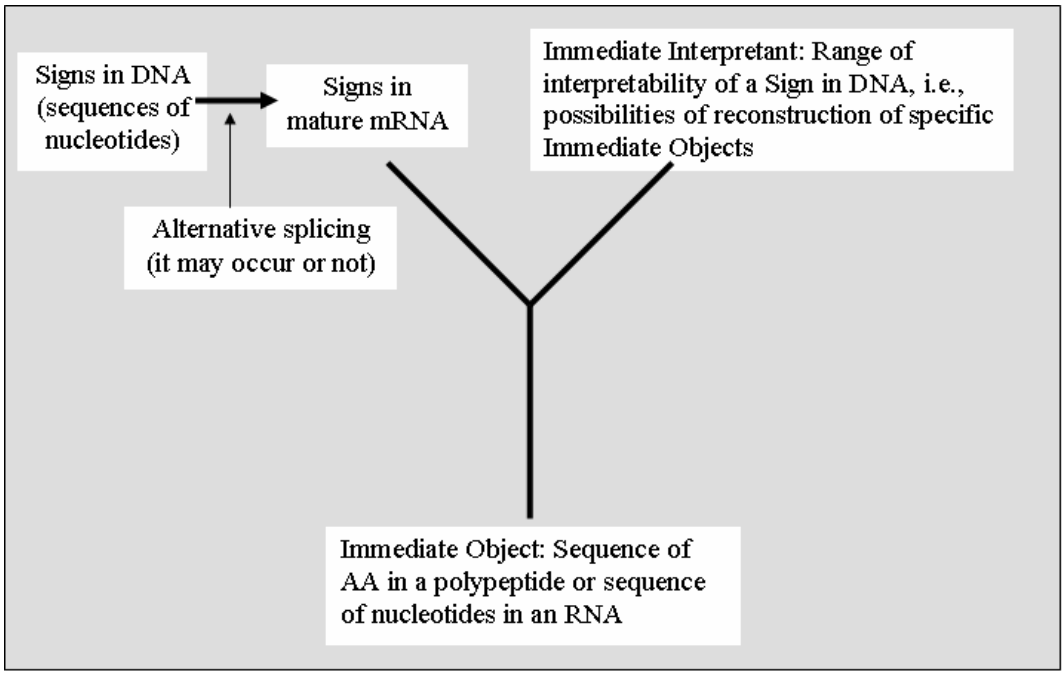

Figure 2: A general semiotic analysis of the gene as a Sign.

In Figure 2, a sequence of nucleotides in DNA is not treated as information in itself, but as the first correlate of information interpreted as semiosis, namely a Sign. Signs in DNA are transcribed into Signs in mature mRNA, with or without the occurrence of alternative splicing, a process through which different patterns of RNA processing lead to a number of different mature RNA molecules, each coding for different, but related proteins (isoforms). If alternative splicing takes place, a Sign in DNA will be used, then, to produce several different Signs in mRNA. The Immediate Object of a gene as a Sign in DNA is the sequence of amino acids or nucleotides represented in it. And as several different Immediate Objects can be represented in DNA (given processes such as alternative splicing), there is a range of interpretability of a Sign in DNA, which amounts to the Immediate Interpretant.

A protein-coding gene, for instance, can only become a Sign in effective action in a cell by standing – in a triadic-dependent relation – for a specific sequence of amino acids (immediate object) through a process of reconstruction of a specific form (Interpretant). What is genetic information in this scheme? It can be understood as the whole process through which a gene acts as a Sign in a given cell, mediating the reconstruction of a specific sequence of amino acids. Information is the triadic-dependent relation *per se*, it is a process, not something to be found in the first correlate of this process, a Sign in DNA. In Signs in DNA, we can only find information *in potency*. When this potential information indeed becomes actual information it is not something contained in isolated signs in DNA, but the very process through which those Signs act.

A Sign is the mediating element in a semiotic process through which a form is communicated from an Object to an Interpretant. This is the reason why we consider the Interpretant here as the reconstruction of a form (habit)

embodied in an Object. ‘Reconstruction’ here amounts to a process by which the form of a protein in a cell generation is communicated through Signs in DNA (in potency) to the form of a protein in the next cell generation. Thus, a regularity obtains in the three-dimensional structure and function of proteins over generations.

The relationship between Signs in DNA and sequences of amino acids in proteins is established by a complex mechanism of interpretation, involving transcription, RNA processing, and translation. Thus, to interpret a string of DNA, more than one interpretative system is required, including, for instance, RNA polymerases, involved in the transcription of DNA into RNA, and ribosomes, involved in the translation of mRNA into proteins. These interpretative systems are parts or subsystems of a cell as a global interpreter, and their actions are subordinated to the latter. The idea that the cell can be seen as a global interpreter to which a series of interpretative subsystems in the genetic information system are subordinated is dramatically reinforced by recent analyses of the functional organization of proteomes. Consider, for instance, that the multi-component cellular systems involved in transcription, RNA processing, and RNA transport do not form a simple linear assembly line, but a complex and extensively coupled network in which signals circulate in a non-linear manner, involving several feedback loops (Maniatis & Reed 2002, Kornblihtt et al. 2004). It is this network structure that makes it possible the coordination of the interpretative subsystems in the genetic information system by the cell. It is clear, then, that we cannot easily move from claims at the cell level to claims at the molecular level while pondering about which system is interpreting genes as Signs. It is becoming increasingly clear through recent advances in the understanding of cell systems that, when a gene is interpreted, the interpretation process is indeed taking place at the cellular level, albeit multi-component molecular subsystems are necessary to this endeavor. This idea that ultimately the whole cell participates in the network necessary for the interpretation that is demanded for the effect of a gene product to take place (Emmeche & Hoffmeyer 1991, Pardini & Guimarães 1992) is further supported by the role of an impressive array of signaling pathways regulating the interpretation of Signs in DNA. As Fogle (2000, p. 19) sums up, “DNA action and function become meaningful in the context of a cellular system. Coding information in the DNA is necessary but insufficient for the operation of living systems.” Accordingly, a Peircean approach to genes and genetic information entails that genetic structures should not be seen in isolation from the larger system by which they are interpreted. From this perspective, the meaning of a gene to its interpreter, the cell, or, to put it differently, the biological meaningfulness of a gene, is found not only in DNA sequences in a chromosome. That there is more to genetic information than just a sequence of nucleotides in DNA means that we will have to include in our models of information the effect of the gene-as-a-sign on the cell or organism, and, in fact, the very role of cellular subsystems as interpreters of strings of DNA, in such a way that they relate Signs to specific Dynamical Objects, proteins that play a function inside the cellular system and have an effect on it or on the organism of which the cell is a part.

From an identification of genetic information with sequential information in DNA, we move in a Peircean framework to its understanding as a triadic-dependent, semiotic process. As a way of stressing the difference between an account of information as a process and more usual explanations about what is information, consider, for instance, Maynard Smith and Szathmáry's (1999, pp. 9-10) argument that information is 'that something' which is conserved throughout a series of changes in the material medium underlying a communication process. According to the model developed above, 'that conserved something' is not information, but rather an invariance in the reconstructed form. Information is rather the process by which a form is conveyed through several different media (Signs) in such a way that an invariance is conserved throughout the process, even though the significant aspects of the object's form are continually reconstructed. Applying this idea in the context of the analysis offered in this section, it is not genetic information that is conserved throughout the different tokens of DNA molecules (different material media) in different organisms and generations, but rather an invariance, that is, a habit or a tendency to build tokens of the same kind based on the Signs available in DNA. This can only be called potential information. Genetic information in itself is the process by which that invariance is conveyed to a new token of a protein, *i.e.*, the whole process through which genes as signs in DNA (entailing the potentiality for genetic information as a process) are irreducibly related to Objects and Interpretants.

Transcription, RNA processing, and protein synthesis can be understood, in semiotic terms, as processes of actualization of potential signs in protein-coding genes. When put into action, a protein-coding gene becomes part of effective semiosis, a triadic-dependent process by means of which the gene as a Sign indicates a given functional product, synthesized after splicing, mRNA edition, or any other complexity involved in the path from a DNA stretch to a protein. This functional product has in turn an effect on the organism in which it is expressed (its Final Interpretant), participating in its adaptive interactions with its surroundings, and, thus, contributing to the presence of that potential gene in the next generation in a high frequency. Notice that we are not postulating any inversion of the central dogma (as if sequences of amino acids in proteins might determine sequences of nucleotides in DNA). We are referring, rather, to the effect of functional proteins on the likelihood that certain genes, certain Signs mediating the process of their synthesis, be present in future generations.

The actualization of a gene depends on boundary conditions established by a higher-level semiotic network, a network of signaling processes that regulate gene expression, ultimately determining the likelihood of transcription of a given gene, or splicing of a given pre-mRNA according to a particular pattern, or chemical modification of a given protein in a manner that modulates its function in a particular way (*e.g.*, by phosphorylation), and so on. A variety of regulatory mechanisms studied in cellular and molecular biology can be thus interpreted as composing a macro-semiotic environment establishing boundary condition that will downwardly determine which potential genes in a string of DNA will be actualized, entering into effective action in a cell.

This shows how several complexities involved in gene expression can be introduced in our analysis: boundary conditions established by this macro-semiotic environment will determine, for instance, which stretch of DNA will be read (e.g., allowing for an analysis of transcription of overlapped or nested genes), which pattern of RNA splicing or RNA editing will be instantiated in order to produce a particular mature mRNA (allowing for the subtleties of alternative RNA splicing or RNA editing to be taken into account), which functional protein will be effectively constructed by the cell (allowing for chemical and/or structural modifications suffered by the primary amino acid sequence of a protein to be considered), and so on. The regulatory influence of the macro-semiotic level, *i.e.*, of the network of signaling processes on interpretative subsystems, and, thus, on transcription, splicing, translation, shows that we have to ultimately consider the whole cell as participating in the network necessary for the actualization of potential genes in DNA. The cellular network of semiotic processes is, in turn, highly responsive to environmental factors, given the semi-open nature of living systems. Accordingly, genes, as potential signs in DNA, are actualized in response to regulatory dispositions arising from a network of signaling pathways that elicit cellular specific responses to other signs arising from a hierarchy of 'contexts', 'environments', or, in our own terms, semiotic levels that can direct gene expression (i.e., establish boundary conditions for the selection of potential genes in DNA), ranging from systems of gene-gene interactions to organisms, and passing through nucleus, cytoplasm, cell, cell surface, extracellular matrix, morphogenetic fields, collective condensations of cells (blastemas), organs, etc. (see, e.g., Hall, 2001). Thus, the cell, as an interpreter, answers to an environmental cue by means of a specific alteration of its internal states, triggered by a whole network of signal transduction culminating in a change at some level of gene regulation (For a semiotic analysis of signal transduction systems, see below. See also Bruni 2003, Queiroz & El-Hani 2006, El-Hani, Arnellos & Queiroz 2007). These relations cannot be understood only in terms of molecular interactions taking place in networks of signal transduction, because this latter process crucially involves semiotic events, as the widespread usage of information talk in modeling and explaining signaling pathways clearly suggests.

This semiotic analysis allows us to offer an interesting account of the 'transmission' of information. It is not effective information that is being communicated when one observe, for instance, 'vertical transmission', from parent to offspring. From the perspective of the model explained above, what is being communicated is only potential information, i.e., the potentiality of a process called information, which can be said, as explained above, to be carried by stretches of DNA. Signs in DNA will only become elements in effective information when interpreted by the cell. Effective information itself cannot be carried from one system to another, but only potential information can be 'carried' by the first correlates of triads, Signs (the vehicles of which, in biological systems, are typically physicochemical entities).

This biosemiotic analysis of the genetic information system leads to the following conclusions:

(i) Genes should be treated as Signs in DNA, which can only have any effect on a cell through a triadic-dependent process (semiosis);

(ii) This process *is* genetic information and involves more than just genes as signs in DNA but also objects and interpretants;

(iii) Genetic information is the process by means of which a form in a dynamical object (a functional protein) is communicated to an interpretant (the reconstruction of a specific sequence of amino acids in a cell) through signs in DNA.

Acknowledgement: João Queiroz and Charbel El-Hani are indebted to the Brazilian National Research Council (CNPq) and the State of Bahia Foundation for Research Support (FAPESB).

Next lecture (Semiotic processes in the immune system): In the next lecture we introduce our model of semiotic processes in the immune system theory.

References

Adami, Christoph (2004). Information theory in molecular biology. *Physics of Life Reviews* 1, 3-22.

Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith and Walter, Peter (2002). *Molecular Biology of the Cell* (4th Ed.). New York: Garland Science.

Atlan, Henry and Koppel, Moshe (1990). The Cellular Computer DNA: Program or Data? *Bulletin of Mathematical Biology* 52 (3), 335-348.

Barbieri, M. (org.) (2007). *Introduction to Biosemiotics*. The new biological synthesis. Dordrecht: Springer.

Black, Douglas L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry* 72, 291-336.

Bruni, Luis Emilio (2003) *A Sign-theoretic Approach to Biotechnology*. Unpublished Ph.D. dissertation. Institute of Molecular Biology, University of Copenhagen.

Cover, Thomas M. and Thomas, Joy A. (1999) Information Theory. In *The MIT Encyclopedia of the Cognitive Sciences*. Robert A. Wilson & Frank C. Keil (eds.), 404-406. Cambridge-MA: MIT Press.

Coyne, Jerry A. (2000). The gene is dead: Long live the gene. *Nature* 408, 26-27.

Crick, Francis H. (1958). On protein synthesis. *Symposium of the Society of Experimental Biology* 12, 138-163.

Debrock, Guy (1996). Information and the metaphysical status of the sign. In *Peirce's Doctrine of Signs - theory, applications, and connections*. V. Colapietro & T. Olshewsky (eds.), 80-89. Berlin, New York: Mouton de Gruyter.

El-Hani, C. N. (2007). Between the cross and the sword: the crisis of the gene concept. *Genetics and Molecular Biology* 30(2): 297-307.

El-Hani, C. N.; Queiroz, J. & Emmeche, C. (2009). *Genes, Information, and Semiosis*. Tartu University Press, Tartu Semiotics Library.

__. 2006. A semiotic analysis of the genetic information system. *Semiotica* 160 (1):1-68.

El-Hani, Charbel Niño; Arnellos, Argyris; and Queiroz, João. 2007. Modeling a Semiotic Process in the Immune System: Signal Transduction in B-cells Activation. *TripleC --Cognition, Communication, Co-operation* 5(2): 24-36.

Emmeche, Claus and Hoffmeyer, Jesper (1991). From language to nature - the semiotic metaphor in biology. *Semiotica* 84 (1/2), 1-42.

Falk, Raphael (1986). What is a gene? *Studies in the History and Philosophy of Science* 17, 133-173.

____. (2000). The gene – A concept in tension. In *The Concept of the Gene in Development and Evolution*, Peter J. Beurton, Raphael Falk and Hans-Jörg Rheinberger (eds.), 317-348. Cambridge: Cambridge University Press.

____. (2001). Can the norm of reaction save the gene concept? In *Thinking about Evolution: Historical, Philosophical and Political Perspectives*, Rama S. Singh, Costas B. Krimbas, Diane B. Paul and John Beatty (eds.), 119-140. New York: Cambridge University Press.

Fetzer, James (2001). *Computers and Cognition: Why Minds are not Machines*. Dordrecht: Kluwer.

Fogle, Thomas (1990). Are genes units of inheritance? *Biology and Philosophy* 5, 349-371.

____. (2000). The Dissolution of Protein Coding Genes. In *The Concept of the Gene in Development and Evolution*, Peter J. Beurton, Raphael Falk and Hans-Jörg Rheinberger (eds.), 3-25. Cambridge: Cambridge University Press.

Gilbert, Walter (1978). Why genes in pieces? *Nature* 271, 501.

Gray, Russell D. (1992). Death of the gene: Developmental systems fight back. In *Trees of Life: Essays in the Philosophy of Biology*, Paul E. Griffiths (ed.), 165-209. Dordrecht: Kluwer.

Griffiths, Anthony J. F.; Gelbart, William M.; Miller, Jeffrey H. & Lewontin, Richard C. (1999). *Modern Genetic Analysis*. New York: W. H. Freeman & Co.

Griffiths, Paul E. (2001). Genetic information: A metaphor in search of a theory. *Philosophy of Science* 68 (3), 394-403.

Hall, Brian K. (2001). The gene is not dead, merely orphaned and seeking a home. *Evolution and Development* 3 (4), 225-228.

Hoffmeyer, Jesper (1996). *Signs of Meaning in the Universe*. Bloomington & Indianapolis: Indiana University Press.

Ideker, Trey; Timothy Galitski and Leroy Hood (2001). A new approach to decoding life: systems biology. *Annual Review of Genomics and Human Genetics* 2, 343-372.

Jablonka, Eva (2002). Information: Its interpretation, its inheritance, and its sharing. *Philosophy of Science* 69, 578-605.

Jablonka, Eva & Lamb, Marion. (2005) *Evolution in Four Dimensions - Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. MIT Press.

Johannsen, Wilhelm (1909). *Elemente der Exakten Erblichkeitslehre*. Jena: Gustav Fischer.

Johansen, Jorgen Dines (1993). *Dialogic Semiosis*. Bloomington and Indianapolis: Indiana University Press.

Kampa, Dione; Cheng, Jill; Kapranov, Philipp; Yamanaka, Mark; Brubaker, Shane; Cawley, Simon; Drenkow, Jorg; Piccolboni, Antonio; Bekiranov, Stefan; Helt, Gregg; Tammana, Hari and Gingeras, Thomas R. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Research* 14, 331-342.

Keller, Evelyn Fox (2000). *The Century of the Gene*. Cambridge-MA: Harvard University Press.

__. (2005). The century beyond the gene. *Journal of Biosciences* 30, 3-10.

Kitcher, Philip (1982). Genes. *British Journal for the Philosophy of Science* 33, 337-359.

Kornblihtt, A. R.; de la Mata, M.; Fededa, J. P.; Muñoz, M. J. and Nogues, G. (2004). Multiple links between transcription and splicing. *RNA* 10: 1489-1498.

Kull, Kalevi (1999). Biosemiotics in the twentieth century: A view from biology. *Semiotica* 127 (1/4), 385-414.

Lewin, Benjamin (2004). *Genes VIII*. Upper Saddle River-NJ: Pearson Education.

Lodish, Harvey; Berk, Arnold; Matsudaira, Paul; Kaiser, Chris A.; Krieger, Monty; Scott, Matthew P.; Zipursky, S. Lawrence and Darnell, James (2003). *Molecular Cell Biology* (5th Ed.). New York: W. H. Freeman & Co.

Magurran, Anne (2000). Backseat drivers, Review of *The Century of the Gene* by E. F. Keller. *New York Times Book Reviews*, 10 December, p. 26.

Maniatis, Tom and Reed, Robin (2002). An extensive network of coupling among gene expression machines. *Nature* 416:499-506.

Maynard Smith, John (2000a).The concept of information in Biology, *Philosophy of Science* 67 (2), 177-194.

____. (2000b). The Cheshire cat's DNA. *The New York Review of Books* 47, 43-46.

Maynard Smith, Jonh & Szathmáry, Eörs (1999). *The Origins of Life: From the Birth of Life to the Origins of Language*. Oxford: Oxford University Press.

Modrek, Barmak and Lee, Christopher (2002). A genomic view of alternative splicing. *Nature Genetics* 30, 13-19.

Moss, Lenny (2001). Deconstructing the gene and reconstructing molecular developmental systems. In *Cycles of Contingency: Developmental Systems and Evolution*, Susan Oyama, Paul E. Griffiths and Russell D. Gray (eds.), 85-97. Cambridge-MA: MIT Press.

__. (2003). *What Genes Can't Do*. Cambridge-MA: MIT Press.

Neumann-Held, Eva (1999). The gene is dead – Long live the gene: Conceptualizing genes the constructionist way. In *Sociobiology and Bioeconomics: The Theory of Evolution in Biological and Economic Thinking*, Peter Koslowski (Orgs.), 105 - 137. Berlin: Springer.

Nijhout, H. Frederik (1990). Metaphors and the role of genes in development. *Bioessays* 12(9):441-446.

Oyama, Susan (2000). *The Ontogeny of Information: Developmental Systems and Evolution*, 2nd Ed. Cambridge: Cambridge University Press. 1st Ed. 1985.

Pardini, Maria Inês Moura Campos and Romeu Cardoso Guimarães (1992). A systemic concept of the gene. *Revista Brasileira de Genética* 15(3), 713-721.

Pitombo, M. A., Almeida, A. M. R. & El-Hani, C. N. (2008). Gene concepts in higher education cell and molecular biology textbooks. *Science Education International* 19(2): 219-234.

Portin, Petter (1993). The concept of the gene: short history and present status. *Quarterly Review of Biology* 56, 173-223.

Queiroz, J. & El-Hani, C. 2006. Towards a multi-level approach to the emergence of meaning in living systems. Acta Biotheoretica 54: 179-206.

Queiroz, J., Emmeche, C. & El-Hani, C. 2005. Information and semiosis in living systems: a semiotic approach. *Semiotics, Evolution, Energy, and Development* 1: 60-90.

Rios, Ricardo Iglesias (2004). O início do fim do gene, resenha de Keller, E. F. O Século do Gene. *Ciência Hoje* 34 (204), 72-73.

Sarkar, Sahotra (1996). Biological information: A skeptical look at some central dogmas of molecular biology. In *The Philosophy and History of Molecular Biology: New Perspectives,* Sahotra Sarkar (ed.), pp. 187-231. Dordrecht: Kluwer.

__. (1998). Forty years under the central dogma. *Trends in Biochemical Sciences* 23, 312-316.

Shannon, Claude E. & Weaver, Warren (1949). *The Mathematical Theory of Communication*. Urbana-IL: University of Illinois Press.

Sorek, Roten; Shamir, Ron and Ast, Gil (2004). How prevalent is functional alternative splicing in the human genome? *Trends in Genetics* 20 (2), 68-71.

Stotz, Karola, Griffiths, Paul E. and Knight, Rob (2004). How biologists conceptualize genes: An empirical study. *Studies in the History and Philosophy of Biological & Biomedical Sciences* 35, 647-673.

Stuart, C. I. J. M. (1985). Bio-informational equivalence. *Journal of Theoretical Biology* 113, 611-636.

Wang, Wen; Zhang, Jianming; Alvarez, Carlos; Llopart, Ana and Long, Manyuan (2000). The origin of the *Jingwei* gene and the complex modular structure of its parental gene, *Yellow Emperor*, in *Drosophila melanogaster*. *Molecular Biology and Evolution*. 17 (9), 1294-1301.

Wilkins, Adams S. 2002. Grappling with developmental complexity. *BioEssays* 24, 1193-1195.

Williams, Nigel (1997). Biologists cut reductionist approach down to size. *Science* 277 (5325), 476-477.