Sci & Educ (2013) 22:345–374
DOI 10.1007/s11191-011-9390-z

How to Understand the Gene in the Twenty-First Century?

Lia Midori Nascimento Meyer · Gilberto Cafezeiro Bomfim · Charbel Niño El-Hani

Published online: 15 September 2011

Abstract It is widely acknowledged in the literature on philosophy of biology and, more recently, among biologists themselves that the gene concept is currently in crisis. This crisis concerns the so-called "classical molecular concept", according to which a gene is a DNA segment encoding one functional product, which can be either a RNA molecule or a polypeptide. In this paper, we first describe three categories of anomalies that challenge this way of understanding genes. Then, we discuss proposals for revising the gene concept so as to accommodate the increasingly known complexity of genomic architecture and dynamics. Our intention is to provide an informative overview of recent proposals concerning how we should conceive of genes, which are probably not very familiar to many science educators and teachers, but can bring relevant contributions to genetics teaching, in particular, to a more critical treatment of genes and their role in living systems.

Keywords Gene · DNA · Unit · Crisis · Genetic teaching

1 Introduction

The gene concept is one of the landmarks in the history of science in the twentieth century. Gelbart (1998) and Keller (2000), for instance, call it "the century of the gene". Moss (2003a), in turn, treats the gene as the central organizing theme of twentieth century biology. At the first decade of the twenty-first century, there are persistent doubts, however, about the prospects that this concept remains as powerful and fruitful as it has been throughout the twentieth century. The perception that the "gene" has become a

L. M. N. Meyer
Department of Biosciences - Campus Itabaiana, Federal University of Sergipe,
Av. Vereador Olímpio Grande, s/n, Centro, Itabaiana, SE ZIP: 49.500-000, Brazil

G. C. Bomfim · C. N. El-Hani (✉)
Institute of Biology, Federal University of Bahia,
Rua Barão de Jeremoabo, s/n, Ondina, Salvador,
BA ZIP: 40170-115, Brazil
e-mail: charbel.elhani@pq.cnpq.br

problematic concept first emerged in the literature on the philosophy of biology (e.g., Kitcher 1982; Burian 1985; Falk 1986; Fogle 1990). Then, doubts and worries about the gene concept began to appear in empirical papers (e.g., Wang et al. 2000; Venter et al. 2001; Kampa et al. 2004), and, finally, by the mid-2000s, they reached the editorials of high-impact scientific journals (e.g., Pearson 2006).

As El-Hani (2007) describes the situation, the gene concept seems to be now between the cross and the sword. By these metaphors, he was trying to show how the gene was, on the one hand, facing claims that we should eliminate it from the biological discourse, denying its usefulness as a scientific term, while, on the other hand, there were several authors trying to save it, by means of diverse proposals of conceptual revision. Keller (2000), for instance, suggested that maybe the time was ripe to forge new words and leave the gene concept aside. However, other philosophers of biology and also practicing scientists took a more optimistic view about the future of the concept, trying to rescue it from a putative dismissal. For instance, Falk, while admitting that the gene was a concept "in tension" (Falk 2000), sought for ways to "save" it (Falk 2001). Hall (2001) was also optimistic, arguing that, despite published obituaries (Gray 1992; Neumann-Held 1999; Keller 2000), the gene was 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. Burian (2004) defended the need of a plurality of gene concepts and argued against the attempt to stabilize gene concepts solely by referring to nucleotide sequences, as a consequence of the phenomena of molecular pleiotropy[1] and molecular epigenesis.[2] Knight (2007) argued that "reports of the death of the gene are greatly exaggerated".

The problems with the gene concept are mostly related to its interpretation as *a stretch of DNA that encodes a functional product, a single polypeptide chain or RNA molecule*. This way of understanding genes has been called the *classical molecular gene concept* (Neumann-Held 1999; Griffiths and Neumann-Held 1999; Stotz et al. 2004). From the 1940s to the 1970s, it seemed that simple and straightforward one-to-one relationships (function = gene = polypeptide = continuous piece of DNA = cistron) were acceptable when one attempted to understand the functioning of the genetic system (Scherrer and Jost 2007a, b). These relationships were captured in a powerful way by the classical molecular gene concept, which treats the gene as an uninterrupted unit in the genome, with a clear beginning and a clear ending, and with a single function being ascribed to its product (and, thus, indirectly, to the gene itself). This is a concept of both a structural and functional unit in the genome, and, indeed, it was by bringing together the structural and functional definitions of the gene that the classical molecular concept showed substantial explanatory, predictive, and heuristic powers. The molecular gene initially had a well-defined structure, with easily determinable beginning and end, a singular function, and an easily understandable mechanics. As a consequence of its success, this concept updated the classical geneticists' interpretation of genes as units, by treating the gene as a discrete region of DNA with simple organization and unitary function. A molecular understanding of the gene was thus superimposed onto the idea of 'unit' found in Mendelian genetics (Fogle 1990).

[1] By "molecular pleiotropy", Burian refers to the production of distinct molecules out of a single putative gene, with a major role being played by the cellular and external environments in determining which protein is produced.

[2] "Molecular epigenesis" concerns the revision of sequence-based information through alteration of molecular conformations or action of noninformational molecules, which plays a major role in development.

Nevertheless, the idea of the gene as a structural and functional unit has been challenged in the last three decades by several anomalies resulting from research in genetics and molecular biology, mostly—but not only—in eukaryotes, in which we find nothing like the tight physical complex linking transcription and translation that we observe in bacteria. Among these anomalies, we find, for instance, alternative splicing,[3] nested genes,[4] overlapping genes,[5] transplicing,[6] mRNA editing,[7] etc. (see, e.g., Falk 1986, 2000; Fogle 1990, 2000; Pardini and Guimarães 1992; Griffiths and Neumann-Held 1999; Keller 2000; Moss 2001, 2003a; El-Hani 2007).

In the present paper, we will discuss proposals for revising the gene concept in the face of its current crisis. Our intention is to provide an informative overview of recent developments concerning our understanding of genes which are probably not very familiar to many science educators and teachers, but can bring relevant contributions to genetics teaching, in particular, to a more critical treatment of genes and their role in living systems. In the next section, we will deal with the crisis of the gene concept in more detail. Then, we will discuss some proposals for revising our views about genes that emerged at the turn of the century, from 1990 to 2003. Then we will consider more recent developments on ongoing efforts to reconceptualize genes (2005–2009). Finally, we will present our concluding remarks, discussing the prospects for transposing these developments to genetics teaching, in the contexts of teacher education, higher education in general, and high school.

2 The Crisis of the Gene Concept

As mentioned above, the so-called crisis of the gene concept (e.g., Falk 1986; Fogle 1990; Keller 2000; Pearson 2006) concerns, in fact, the situation faced by a particular way of understanding genes, which was—and, generally speaking, still is—widely accepted among biologists, in school science (either at basic or higher education), and among the public opinion, the *classical molecular gene concept* (Neumann-Held 1999; Griffiths and Neumann-Held 1999; Stotz et al. 2004). In this concept, genes could be convincingly treated as both functional and structural units, and, with the introduction of an informational vocabulary in molecular biology and genetics (see Kay 2000), genes were also regarded as informational units, leading to what has been called the informational conception of the gene (Stotz et al. 2004), another very popular notion in textbooks, the media, and public opinion.

Nevertheless, the idea of the gene as a structural, functional, and informational unit has been challenged in the last three decades by several anomalies resulting from research in genetics and molecular biology. We can classify these anomalies in three kinds, all related to counterevidence for a unitary relationship between genes, gene products, and gene

[3] In alternative splicing, a pre-mRNA molecule is processed—in particular, spliced—in a diversity of manners, so that different combinations of exons emerge in the mature mRNA. In this manner, several distinct mRNAs and, thus, polypeptides can be obtained from the same DNA sequence. In the case of DSCAM in *Drosophila melanogaster*, for instance, alternative splicing can lead to ca. 38,016 protein products (Celotto and Graveley 2001).

[4] A gene is said to be nested when it is entirely located inside another gene.

[5] Genes are said to be overlapping when they share DNA sequences.

[6] In transplicing, mature RNA is formed during processing from RNA transcripts of DNA regions from different chromosomes.

[7] mRNA editing is an alteration of mRNA nucleotides during processing, so that there can be a lack of correspondence between nucleotide sequences in mature mRNA and nucleotide sequences in DNA.

function (Pardini and Guimarães 1992; El-Hani 2007): (i) *one-to-many* correspondences between DNA segments and RNAs/polypeptides (as, for instance, in alternative splicing, see Black 2003; Graveley 2001; Ast 2004); (ii) *many-to-one* correspondences between DNA segments and RNAs/polypeptides (as in genomic rearrangements, such as those that take place in the generation of diversity in lymphocyte antigen receptors in the immune system,[8] see Cooper and Alder 2006; Murre 2007); and (iii) *lack of correspondence* between DNA segments and RNAs/polypeptides (as we see, for example, in mRNA editing, see Hanson 1996; Lev-Maor et al. 2007).

These kinds of anomalies have been accumulating since the 1970s and came to be regarded as being so serious that some authors even proposed that the gene concept needed to be abandoned (Portin 1993; Gelbart 1998; Keller 2000). More frequently, however, we find attempts to save it, by means of a series of conceptual revisions aiming at solving the problems generated by the anomalies challenging the classical molecular concept. Our intention in this paper is precisely to discuss these proposals. Some of them reflect the transformation suffered by our understanding of genes as a consequence of investigations conducted in the post-genomic era. Gerstein et al. (2007), for instance, put forward a reformulation of the gene concept that is strongly grounded on the Human Genome Project (HGP) and the ENCODE (*ENCyclopedia Of DNA Elements*) Project, which added one extra layer of complexity to the understanding of genome architecture and function, bringing additional difficulties to the classical molecular concept. Consider, for instance, the existence of a diversity of non protein coding RNAs (ncRNAs), some of them with functions in gene regulation and RNA processing; transcribed pseudogenes, with direct or indirect regulatory function[9]; and the fact that many ncRNAs and transcribed pseudogenes originate from genome regions previously regarded as "junk DNA", showing how this designation is entirely inadequate, since we are dealing with transcriptionally active DNA segments.

In this paper, our focus lies in developments regarding the gene concept that took place in the last 20 years. First, we will consider developments that mostly happened in the 1990s, and then we will move on to consider more recent proposals, put forward in the last 5 years.

3 Rethinking the Gene Concept at the Turn of the Twenty-First Century (1990–2003)

We will consider here the following proposals for revision or analysis of the gene concept: The process molecular gene concept (Griffiths and Neumann-Held 1999; Neumann-Held 2001); the distinction between gene-P and gene-D (Moss 2001, 2003a, b); the treatment of genes as sets of domains in DNA (Fogle 1990, 2000); and the systemic gene concept (Pardini and Guimarães 1992).

[8] The generation of the diverse antigen receptors found in lymphocytes, and, consequently, of antibody specificity depends on a combinatorial set of genomic rearrangements between different DNA segments called variable segments, constant segments, and diversity and joining segments.

[9] Pseudogenes are genomic DNA sequences which are derived from, and similar to protein-coding genes, but show signs diagnostic of protein-coding deficiency, such as frameshifts and premature stop codons, and are usually non-functional. Transcribed pseudogenes are copies of protein-coding genes that have accumulated such indicators of coding sequence decay, but are still transcribed, and could be potentially functional in the regulation of gene transcription (Khachane and Harrison 2009).

3.1 Process Molecular Genes

The "process molecular gene concept" treats genes not as "bare DNA", but rather as the whole molecular process underlying the capacity to express a particular product (a polypeptide or RNA). According to it, "… 'gene' denotes the recurring process that leads to the temporally and spatially regulated expression of a particular polypeptide product" (Griffiths and Neumann-Held 1999, p. 659). That is, this way of understanding genes is focused on how DNA sequences are used in the process of producing polypeptides or RNAs. This process involves not only DNA, but all the entities that participate in the synthesis of RNAs or polypeptides. In this manner, different epigenetic conditions which can affect gene expression are built into the gene.

To understand this move, we can consider Epp's (1997) proposal that we need to separate two distinct concepts related to genes, namely, a specification of what is a gene and an indication of how it is used. But, while Epp claims that the term "gene" points to the former, Griffiths and Neumann-Held argue that it rather points to the latter. They take, thus, a more process- than substance-oriented view on genes, stressing how they are used, rather than what they are as physical entities. They justify this option by emphasizing that "the concept of the gene has always been intimately linked to how genes are used in development" (Griffiths and Neumann-Held 1999, p. 658). In the process molecular gene, the focus lies on gene function, and a description of what is a gene as a physical entity is taken as a necessary but not sufficient condition to understand it, given the context-dependence of the function played by a given stretch of DNA.

Some problems faced by the gene concept may also support the move towards a process-oriented interpretation. Consider, for instance, the case of regions in overlapping genes which are inside the open reading frame (ORF)[10] of a given gene while also function as promoters for other overlapped genes. Such situations suggest that whether a region in DNA is part of a particular gene or not depends on which function it performs relatively to that gene. It follows that functional descriptions of regions in DNA, such as "gene", "promoter", "enhancer", cannot be explained merely in terms of structural descriptions, such as those of specific sequences in DNA.

The process nature of this concept arguably makes it possible to accommodate anomalies which the classical molecular gene has difficulties in facing, such as alternative splicing or mRNA editing. These phenomena are simply included in the gene, when interpreted as a process, dissolving the anomalies at stake.

Moss (2001) points out, however, a number of potentially troublesome consequences of the process molecular gene. First, it substantially increases the number of genes in eukaryotes, due to the great number of polypeptides generated by alternative splicing. Second, it demands that we include in genes the multimolecular systems associated with transcription and splicing, making the process molecular gene jump to a higher level in the biological hierarchy. Third, it is hard to individuate process genes, given the context-dependence of gene expression.

These problems, albeit serious, are not necessarily fatal. Concerning the first problem, we can pose the following question: What if an approach that we ultimately accept as appropriate to ascribe meaning to the gene concept has, as one of its consequences, an increase in the number of genes? It does not make much sense to give up an approach,

[10] An open reading frame is the DNA or RNA sequence located between the initiation codon, where the codons that will be translated into amino acids in protein synthesis begin to be read, and the termination codon, where protein synthesis comes to an end.

if the scientific community eventually regards it as adequate, just because of an increase in the number of genes to be counted. Moreover, to increase the number of genes is far from being necessarily a problem, as we can see by considering the *N* value paradox (Claverie 2001), i.e., the largest size of the proteome when compared to the genome. For example, the number of human genes estimated by the HGP was far below previous estimates, which ranged from 50,000 to more than 140,000 genes. While the public effort reported 30,000–40,000 (International Human Genome Sequencing Consortium 2001), the private endeavor identified 26,000–38,000 transcriptional units (Venter et al. 2001). These numbers are rather close, for instance, to the gene content of *Drosophila melanogaster* (14,000 genes), *Caenorhabditis elegans* (ca. 19,000), and *Arabidopsis thaliana* (25,498). The number of human proteins, however, can be estimated in 90,000 (Magen and Ast 2005). This is the basis for the paradox: if each gene coded for a single protein, how could we explain, say, that the human proteome is about five times bigger than that of the fruit fly? As the process molecular gene includes the mechanisms leading to proteome expansion, such as alternative splicing, it could accommodate this paradoxical relationship between genome and proteome. Therefore, the first problem pointed out by Moss can even end up being an advantage of the process molecular concept.

Regarding the second problem, we can argue that biological concepts are usually defined from an organismic perspective, in which concepts at a given hierarchical level, say, biomolecules, are related to concepts at other levels, such as cells or organisms. This means that genes are also dependent on higher levels at the biological hierarchy, as a natural consequence of the complexity of living systems. Maybe this alleviates the problem posed by Moss, but we should admit that the jump of the process molecular gene to a higher level in the biological hierarchy is indeed a counterintuitive aspect of this way of treating genes. We cannot forget, however, that scientific advances are often counterintuitive.

Finally, with regard to the third problem, the difficulty of individuating genes can be treated as a characteristic rather than a problem of process-oriented approaches. Given their context-dependence and lack of well-defined boundaries, any processes are harder to individuate than entities. Nevertheless, just as in the case of the first problem, the fact that we will need to deal with genes that are harder to individuate can be a weak reason for abandoning a process approach to genes, if we ever come to conclude that this is the most appropriate way of understanding them. Perhaps, it is also an advantage of this concept that it makes genes harder to individuate, since in this manner it might block atomistic or extremely reductionist interpretations of genome architecture and dynamics.

3.2 Gene-P and Gene-D

Moss (2001, 2003a, b) developed an analysis of the meanings ascribed to genes, which led him to demarcate between two concepts, gene-P and gene-D. Gene-P amounts to the gene as a determinant of phenotypes or phenotypic differences. As Moss (2003a, p. 45) writes:

> Genes for phenotypes, i.e. Genes-P, can be found, generally […] where some deviation from a normal sequence results with some predictability in a phenotypic difference.

It is the "… expression of a kind of instrumental preformationism" (Moss 2001, p. 87), i.e., it is an instrumental concept, not accompanied by any hypothesis of correspondence to reality, and this is what makes it acceptable the simplifying assumption of a preformationist determinism (as if the trait was already contained in the gene, albeit in potency).

This concept is useful to perform a number of relevant tasks in genetics, such as pedigree analysis or genetic improvement by controlled crossing methods.

When we refer to gene-P, we speak of it as if it causes, by itself, a phenotype, as in the commonly used concept of a "gene for" (Kendler 2005), which became rather central in "gene talk" (Keller 2000), both in science and in general society, and can be captured by the statement that "*X* is a gene for *Y*," in which *X* is a particular gene in the human genome and *Y* is a human disorder or trait. For instance, when we speak of the gene for blue eyes, we speak of the genes as if it determined the trait "blue eyes". A "gene for" a trait is shorthand for "a locus in which sequence variation causes a difference in phenotype, all other things being equal" (Knight 2007, p. 298, citing Williams 1966). However, there are many cases in which the "gene for" the trait has no material counterpart, since a whole series of alleles affecting the normal functioning of the biochemical processes involved in the development of the trait—for instance, the biochemical pathways that lead to the synthesis of eye pigments—can result in that trait. Thus, the concept "gene for blue eyes"—an example used by Moss himself—corresponds, in fact, to a disjunction of alleles that can be responsible for a decrease of pigmentation in the iris. And, certainly, this disjunction is a logical expression, not a material entity to which the concept can be said to refer. This does not deny, however, the usefulness of the concept: to understand the results of a crossing between a brown-eyed father and a blue-eyed mother, we can readily use pedigree analysis, accompanied by the apt simplification of assuming that there are genes that determine the presence of brown or blue eyes. We can say, thus, that gene-P is a useful fiction, an instrumental concept that has predictive power and plays an important role in some explanatory games of genetics and molecular biology.

Gene-D, in turn, amounts to the gene as a developmental resource, which is, in itself, indeterminate with respect to the phenotype (Moss 2003a, p. 46).[11] It is typically conceived in the scientific discourse in a more realist manner, as a real entity defined by some molecular sequence in DNA. It is conceived as a transcription unit, providing molecular templates for the synthesis of gene products—i.e., genes-D are usually conceived as in the classical molecular concept. Nevertheless, genes-D typically does not determine by itself phenotypic traits; they are rather developmental resources involved in the construction of traits, alongside with other equally important resources, such as epigenetic and environmental factors.[12] It is by considering its role in development that we can ascribe to it an important explanatory role in genetics, molecular biology, developmental biology, etc.

Moss argues that genes can be productively conceived in these two different ways, *but nothing good results from their conflation* (Moss 2001, p. 85). Gene-P and gene-D are distinct concepts, with different applications, in different explanatory games. For Moss, there is nothing that is, at the same time, a gene-D and a gene-P, i.e., the same sequence cannot count as both gene-P and gene-D. However, these two concepts have been united by means of the rhetorical glue of the gene-as-text metaphor (Moss 2003a, p. 184).

Indeed, in textbooks, at different educational levels, the conflation of these concepts is very common (Pitombo et al. 2008; Santos and El-Hani 2009; Santos et al. in press).

[11] It is important to avoid losing from sight that the distinction between gene-P and gene-D is not identical to the distinction between classical and molecular genes. Molecular entities can be treated as genes-P, as when we speak of the gene for cystic fibrosis (see Moss 2001, 2003a).

[12] Moss appeals to a central notion in Susan Oyama's "developmental systems theory" or "perspective", namely, causal parity between genes and other developmental resources (See Oyama [1985]2000; Oyama et al. 2001). This perspective highlights the missing element in deterministic accounts of the genotype-phenotype relationship, namely, development, in which genes, organisms, and environments interact with each other in such a way that each is both cause and effect in a complex way (Lewontin 1983, 2000).

Furthermore, important social consequences follow from this conflation, since it is one of the main sources of genetic determinism: if one learns about genes that determine traits with no clear explanation of the nature of the model at stake, the instrumental character of the concept, and the kind of explanatory game in which it is appropriate to use it, and then learns, afterwards, about genes in DNA in a rather realist manner, it will straightforwardly follow that those genes found in DNA are so determinative of phenotypic traits as those genes used, say, to solve pedigree problems. When these two concepts are conflated, DNA becomes a developmental program, containing information for the construction of organisms stored in its nucleotide sequences, which needs only to be correctly read for development to take place. Other factors, such as epigenetic and environmental ones, can be, at most, triggers of developmental processes, while DNA is assumed to be both the information reservoir to all development, and an all-powerful controlling molecule. Genes become the major or even single causal determinants of the development of phenotypic traits in quite general terms, even when one considers complex traits, such as sexual orientation, intelligence, or aggression.

This genetic determinist view has become in the last three decades a major feature of gene talk in society, the media, and school science. Therefore, we consider as a very important task for genetics teaching to put this view under a critical lens. Moss' distinction between gene-P and gene-D is quite relevant in an attempt to do so, as it brings to the fore the demarcation of these concepts and their domain of application, as well as the epistemological recognition of their distinct nature (more instrumentalist or realist).

However, we need to consider, also, the fact that Moss' work has been the target of important criticisms, as we find, for instance, in Knight (2007). This author asks whether the distinction between two gene concepts, gene-P and gene-D, accurately reflects the multifarious uses of the term "gene" in biology, and, moreover, whether the diversity of gene concepts indeed hindered biological research. Knight intends to show that the claims that there are precisely two gene concepts and confusion between them has misled biologists do not withstand close scrutiny. He calls attention to the widely acknowledged and documented fact that biologists use many working definitions of genes, and it is even the case that no strict definition of the term may be available, but just a series of generalizations based on well-studied examples of "genes", as conceived by the biological community (Waters 1994). The differences between these working definitions do not preclude—Knight argues—biologists from being able to communicate with one another about genetic phenomena. He opposes, then, this state of affairs in the scientific communities dealing with genes with Moss' claim that there are two concepts at work.

Knight (2007) goes further, however, arguing that the complete incompatibility between gene-P and gene-D—so that no sequence could count as both—results from restrictions that Moss himself introduced, rather than from the ways they are used in practice by biologists. He has two restrictions in mind: on the one hand, the gene-P, which is, in his view, essentially the allele, is complemented by the restriction that all alleles with the same phenotypic effect count as the same gene-P; on the other hand, the gene-D, which is essentially the classical molecular gene, comes together, in Moss' arguments, with the restriction that two sequences with even a single nucleotide difference cannot count as the same gene-D. Accordingly, Knight proposes that the definitions put forward by Moss can be relaxed. For instance, the definition of gene-D can be relaxed by referring "… to a set of sequences that produce the same molecular product (perhaps at the same level and with the same regulation)", with the consequence that "… in most cases the gene-D and the gene-P become one and the same" (Knight 2007, p. 302).

With regard to our use of the distinction between gene-P and gene-D, the fact that there are more ways of conceiving of genes than the two concepts discriminated by Moss does not affect our argument that he indeed pointed out to an important distinction. Indeed, we do think that there are more concepts related to genes being used in the scientific community than gene-P and gene-D, or, as Knight (2007, p. 300) puts it, genes-P/genes-D cannot do full justice to the range of gene concepts. This does not mean, however, that the distinction between gene-P and gene-D cannot have its importance preserved, both in scientific research and science education, among a broader range of conceptual variation. Moreover, even if in a number of cases the same sequence could count as both gene-P and gene-D, if we eliminated the restrictions employed by Moss, we can still see a source of confusion in the mere conflation between these two concepts whenever the simple genotype-phenotype relationship assumed in gene-P, for the sake of its predictive role, is transferred to gene-D, making us lose from sight that in many cases genes do not determine phenotypes, particularly above the level of protein primary structures.

The prospect that the epistemic practices of scientific communities can counteract possible negative consequences of the conflation between these two concepts also does not deny the potential harm resulting from the confusion between gene-P and gene-D at students' learning, practitioners' teaching, and social discourses about genes. It is evident, for instance, that in many cases scientists have a clear view about the way they are conceptualizing genes as being "for a trait", connecting sequence variation with phenotypic differences, as Knight illustrates with Riordan and colleagues' (1989) work on the cystic fibrosis locus. However, we cannot conclude from such examples that problematic conflations between genes-P and genes-D are not made by scientists and, with increasing likelihood, by students, teachers, and lay people.

As an example, we can compare the report of a *correlation* between polymorphic markers in the X_q28 region of chromosome X and male homosexual orientation in a group of 40 families in which there were two homosexual brothers in Hamer et al. (1993) and the simple and straightforward statements that there would be a "gene for" homosexuality or even a "gay gene" in popular science texts (Hamer and Copeland 1994) and scientific papers (Turner 1995), which end up perpetrating the fallacy of confusing correlation with *causation*. With regard to gene talk, we cannot neglect that it was the alleged "gay gene" that made its way to newspapers in a noticeable manner. A clear conflation between gene-P—which embodies just a correlation between sequence variation and phenotypic differences—and gene-D—playing a causal role in the etiology of male homosexual orientation—is clearly at work in this example. This shows that scientists do not have always such a clear view about statements correlating DNA sequences and traits, all other things being equal, suggesting that, if Moss is just restating the idea that the phenotypic effect of a sequence depends on the context of expression through his distinction, as Knight (2007, p. 299) argues, the distinction between gene-P and gene-D still has a role to play in scientific investigation and science education. If scientists may have a hard time dealing with the conflation between gene-P and gene-D, as suggested by the example above, what to say, then, of students, teachers and lay people with much less experience in dealing with the complex and subtle nature of genetic concepts and models? We can conclude, thus, that the importance of the distinction between gene-P and gene-D, as proposed by Moss, can be maintained, both in science practice and science education, despite criticisms such as those found in Knight (2007), but with more caveats than we observe in Moss' original works.

3.3 Genes as Sets of Domains in DNA

In Fogle's (1990, 2000) proposal, genes are treated as sets of domains in DNA. He argues that we should abandon the classical unit concept and recognize that a gene is constructed from an assemblage of embedded, tandem, and overlapping domains in DNA, if we want to take in due account the complexity and diversity of genomic architecture and dynamics. The basic idea is to employ set structure in order to leave aside the need to find a single unit in a region of genetic information. He argues that a gene looks like "… a collection of component entities that together define its structure and influence the phenotype" (Fogle 1990, p. 367). That is why, in his view, a gene is best described as a set of "domains", which are, in turn, nucleotide sequences that can be distinguished from each other on the basis of their structural properties and/or activities: exons, introns, promoters, enhancers, operators, leader and trailer sequences, etc. Domains can be combined in a variety of ways to form sets, or, as Fogle (1990, 2000) calls them, *Domain Sets for Active Transcription* (DSAT). It is not necessary, however, to specify all domains influencing expression for nominative or heuristic purposes, but just "sufficient itemization to prescribe a set that has *communal agreement*" (Fogle 1990, p. 368. Emphasis added).[13]

When understood in this manner, genes are not in DNA anymore. Only domains are found in DNA. A single domain can be part of more than one gene, so that no unit can be found in DNA that corresponds to the gene. A positive aspect of this proposal is that it accommodates anomalies such as overlapping and nested genes, by breaking with the idea of the gene as a structural unit in DNA.

Fogle's proposal implies that the number of DSATs will be much higher than current gene counts. But, just as in the case of the process molecular gene, rather than a problem, this can end up being an advantage of this account, given that this increase may offer a more accurate relationship between the size of the genome and the size of the proteome.

For El-Hani et al. (2009), despite Fogle's negative appraisal of instrumentalist views about genes (Fogle 1990), his proposal may indeed suggest an instrumental approach: DSATs seem to be constructs established by agreement in a community of researchers. To be sure, there is a clear realist side to Fogle's proposal: he treats domains in a realist way, assuming hypotheses of correspondence between concepts like exons or enhancers and actual structures in DNA. But, when building sets of domains, we seem to be, to use Falk's (1986, p. 164) words, superimposing "structure and organization upon the aggregates". As epistemic objects (Rheinberger 2000), DSATs would be built by communities of researchers that reach a communal agreement regarding how to manipulate empirical data and organize theoretical knowledge in their fields.

[13] In order to put Fogle's proposal to work, one has to deal with the rather loose and sometimes confusing usage of terminology in molecular genetics, resulting from the expansion of the zoo of instrumentally formulated genetic entities in the last three decades (Falk 1986). Fogle demands that domains should be clearly specified by structure and/or activity. Therefore, a first task is to build a formal system to designate and describe domains in DNA. This can be done through gene ontology efforts, as developed, for instance, by the Gene Ontology Consortium (Ashburner et al. 2000) and the ENCODE Project (The ENCODE Project Consortium 2007). A second development is suggested, however, by Fogle's ideas, namely, the establishment of formal procedures for the combinations of domains in genes, taking in due account, as far as possible, the practices currently used by the communities of geneticists and molecular biologists. After all, they would ultimately have to make use of the libraries of domains and formal rules of combination resulting from such an effort.

3.4 The Systemic Gene Concept

Pardini and Guimarães (1992) also argue against the idea of genes as units and propose, instead, a systemic concept of the gene, according to which

> the gene is a combination of (one or more) nucleic acid (DNA or RNA) sequences, defined by the system (the whole cell, interacting with the environment, or the environment alone, in subcellular or pre-cellular systems), that corresponds to a product (RNA or polypeptide) (p. 717).

They also distinguish between two kinds of correspondence between the inherited nucleic acid segments and their products, and, consequently, two types of genes: a univocal type, predominant in bacteria, in which a one-to-one correspondence between nucleic acid sequences and gene products holds, and only gene regulation depends on the system in which those sequences are embedded; and a multivocal type, more common in eukaryotes, in which the defining role of the system increases, affecting both the structure and function of the gene.

In both Fogle's DSATs and Pardini and Guimarães' systemic multivocal genes, it is tempting to conclude—if we assume a realist position—that the combinations of nucleotide sequences corresponding to genes might be found in mature RNA molecules, after processing. As we will discuss later, this proposal was indeed put forward in an explicit manner by Scherrer and Jost (2007a, b).

In a subsequent work, Guimarães and Moreira (2000) argue that the meaning of a DNA segment is relative, depending on the expression system in which it is embedded. Consequently, its meaning can be plural: the multivocal nature of genes, particularly in eukaryotes, is related to this context-dependence of gene expression.

Differently from process molecular genes, the systemic gene concept alludes to the process which specifies or demarcates the gene, as an adequate combination of (one or more) genomic sequences corresponding to a product, but does not include this process itself into the gene.

4 More Recent Developments About the Gene Concept (2005–2009)

The genome projects and, in particular, the HGP brought about what has been called the "post-genomic era". In the HGP, by itself, we can see hesitation with regard to the understanding of genes. In a glossary found in a website associated with this project, the gene is defined as "the fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (i.e., a protein or RNA molecule)".[14] In these post-genomic times, this can be regarded as a rather conservative definition, basically corresponding to what has been called the classical molecular concept. It treats genes as units of inheritance, structure, and function. This definition is hard to sustain in the face of the challenges that led to the crisis of the gene concept.

Despite this definition, found in a glossary associated with the HGP, in one of the papers that presented the draft of the human genome sequence, we find a different understanding about the gene. Venter et al. (2001) conceives the gene as "a locus of cotranscribed exons", based on the consideration that "a single gene may give rise to multiple transcripts, and thus multiple distinct proteins with multiple functions, by means of alternative splicing and

[14] http://www.ornl.gov/sci/techresources/Human_Genome/glossary/glossary_g.shtml. Accessed at August 8th 2011.

alternative transcription initiation and termination sites" (p. 1317). On the one hand, this is undeniably an important advance in relation to the conservative definition found in the glossary quoted above, to the extent that it moves away from the classical molecular concept and manages to accommodate phenomena such as alternative splicing. But, on the other hand, this definition still faces problems. It assumes the inclusion of exons corresponding to trailer sequences in the gene, even though there are splicing patterns that result in transcripts that differ from each other by the presence or absence of such trailer sequences. This problem can be solved by considering, as suggested by Fogle (1990), the possibility of excluding both leader and trailer sequences from the gene, maintaining in it only coding exons. Nevertheless, there are still remaining difficulties: alternative splicing can also affect the size of the coding exons. Thus, the idea of taking exons as units of structure in the genome—which underlies proposals such as Venter and colleagues'—is also challenged by alternative splicing. Even if we consider this case of alternative splicing as an exception and, in this manner, try to keep the definition put forward by Venter and colleagues, we would still have to face a host of other challenges, such as mRNA editing, to quote just one example.

Recent advances in molecular biology, genomics, and proteomics made it more and more difficult to conceive of genes as units. It is now quite clear that biological information operates at multiple hierarchical levels, in which complex networks of interactions between components are the rule, and, consequently, the understanding of the dynamics and even the structure of genes demands that they are located in complex informational networks and pathways (Ideker et al. 2001). More recently, new conceptual advances have been made by several authors, which have suggested novel manners of thinking about genes, such as the proposal of a new gene definition based on the results of the ENCODE project (Gerstein et al. 2007); the defense of a shift of the ontological status of the gene from that of a physical entity in DNA to that of a process, as was previously proposed by Griffiths and Neumann-Held (1999) and Neumann-Held (2001); and the introduction of new ways of talking about genes (Keller and Harel 2007; Scherrer and Jost 2007a, b).

4.1 The ENCODE Definition of Gene

The *ENCyclopedia of DNA Elements* (ENCODE) project, an international consortium of scientists trying to identify the functional elements in the human genome sequence, has had a significant impact on our understanding about genes and genomes.[15] Among its contributions, we find a new definition of gene, according to which a nucleotide sequence should satisfy three conditions in order to be a gene: "(1) A gene is a genomic sequence (DNA or RNA) directly encoding functional products, either RNA or protein. (2) In the case that there are several functional products sharing overlapping regions, one takes the *union* of all overlapping genomic sequences coding for them. (3) This union must be *coherent*—i.e., done separately for final protein and RNA products—but does not require that all products necessarily share a common subsequence" (Gerstein et al. 2007, pp. 676–677, emphasis in the original). Based on these conditions, they concisely define the gene as *a union of genomic sequences encoding a coherent set of potentially overlapping functional products* (Gerstein et al. 2007, p. 677, emphasis in the original).

As a consequence of this definition, different functional products of the same class (proteins or RNAs) that overlap in their usage of the same primary DNA sequences are

[15] The ENCODE database can be reached at http://www.genome.gov/10005107#4. The participants of the ENCODE can be found at http://www.genome.gov/26525220. See also (The ENCODE Project Consortium 2007).

combined in the same gene. Thus, an important characteristic of this definition is that the focus is on the products, and, consequently, there is no *one-to-one* relationship between a coding sequence at the DNA level and a functional product. This can be regarded, then, as a promising definition, since it may accommodate several anomalies challenging the unitary relationship between genes, gene products, and gene function that is embedded in the classical molecular concept. For example, the products of alternative splicing are taken to be products of a single gene, since they share sequences in common. The different protein products originating from a single polycistronic pre-mRNA[16] are not regarded as being derived from a single gene, if they do not share any block of nucleotide sequence, and so forth.

Moreover, this gene definition can accommodate one of the most surprising findings of genomic research in the last years, namely, that even though a tiny portion of the human genome consists of open reading frames (ORFs), the majority of the DNA is transcribed into non-protein-coding RNA (ncRNA), whose functions remain largely unknown, but have been increasingly well understood in recent times (e.g., Uney and Lightman 2006; Wu et al. 2006; Carthew 2006; Karres et al. 2007; Valadi et al. 2007; Chien 2007; Niwa and Slack 2007; Zhang et al. 2007; Hendrickson et al. 2009; Hilgers et al. 2010). Li et al. (2011), for instance, found more than 10,000 exonic sites in which RNA sequences do not match DNA sequences, when they compared RNA sequences from human B cells obtained from 27 individuals with the corresponding DNA sequences from the same individuals.

However, if we consider some consequences of the ENCODE definition of the gene, such as those discussed by Smith and Adkison (2010), we will be able to point out some problems. First, untranslated (regulatory) regions (UTRs) and distant gene-associated regulator sequences (translated or not) are not included in the gene; second, the definition does not put limitations on the exon loci that can be combined to form a given product, so that exons can be in different strands of a chromosome, or even on different chromosomes (in the case of transplicing), and still be part of the same gene; third, transcriptionally active regions (TARs) are regarded as putative genes, demanding further investigation.

Some of these consequences contradict criteria for defining genes put forward by Gerstein and colleagues themselves. They mention five criteria that a gene definition should satisfy: (1) It must be *backward compatible* with the past, in the sense that something that used to be called a gene should remain a gene; (2) It must be *organism-independent*, being valid for all biological diversity; (3) It should be a statement of a *simple idea*, not a list of various mechanisms and exceptions; (4) It should be *practical enough* so that one can readily enumerate how many genes there are in a given genome; (5) It should be *compatible with other biological nomenclature* that makes use of the idea of a digital gene. What the consequences above show is that Gerstein et al.'s definition faces difficulties in satisfying the third requirement, to the extent that it gives room to a number of exceptions.

Scherrer and Jost (2007b) also offer critical remarks on the ENCODE definition of the gene. They argue that the gene concept proposed by the project is a hybrid of two aspects related to genetic systems, codification and function, leaving aside a third fundamental aspect, regulation, which mediates between the two former aspects. That is, to the extent that the very goal of the ENCODE project was to systematically describe and classify transcripts, the regulome[17] ended up being neglected by the gene definition associated with it. Thus, it also does not satisfy the fifth criterion stated by its very authors.

[16] A polycistronic pre-mRNA is a single long transcript coding for the syntheses of more than one protein.

[17] The regulome is the complete set of components involved with regulation in a cell.

The difficulties faced by Gerstein and colleagues' proposal when we consider the criteria put forward by themselves are, in our view, a sign of the constraints found whenever we try to look for a universal definition for an overarching biological concept, such as the gene. We will come back to this point in the concluding remarks of this paper.

4.2 Genes as Processes

In two relatively recent works, we find movements in the same direction taken by Griffiths and Neumann-Held in their process molecular gene concept, namely, to suggest that genes can be treated as processes constructed by the cell by using nucleotide sequences, rather than as physical entities at the DNA level.

Even though this was not fully developed in Keller (2005), we interpret her as moving in this direction 6 years ago, when taking a more optimistic view about the future of the gene, in relation to her previous arguments for the abandonment of the gene concept (Keller 2000). Even though she was still advocating that we should move on to a "century beyond the gene", arguing that the twenty-first century will be the century of *genetic systems*, rather than of the gene, she considered then that the gene could yet survive, but only if radically resignified (Keller 2005). The challenges posed by biological complexity demanded "new ways of talking" about genes and, more than that, about living systems in general (Keller 2005, pp. 8–9). It is true that 2 years later (in Keller and Harel 2007), we find no place for the gene in the new vocabulary proposed by this philosopher. Nevertheless, it is interesting to see what she suggested in 2005 as a possibility of reframing our ways of talking and thinking about genes, in a process-oriented direction.

In her suggestion of a profound change in our understanding of living beings, Keller proposed that we should shift our focus from entities, or living systems' components seen in isolation, towards a comprehension of their interaction processes in rather complex networks. This would be an important change, in both theoretical and methodological terms, since it would not be the case that we should only understand the structure and function of the parts constituting biological systems, but we ought to know about how they are connected with each other, how they mutually operate, how they interact in organization levels with diverse degrees of complexity and, moreover, nested inside each other.

Keller considers that, in order to build this new way of understanding biological systems, we need to overcome ingrained habits of thought and speech that give priority to the systems' parts, rather than to the system as a whole. That is, more than new methods (no matter if microarrays, DNA chips, high-throughput sequencing methods, and other techniques about which many scientists are understandably so excited these days), what we will need to get a truly systemic biology is to change our thinking and language.[18]

Among the parts of the living systems, none received more emphasis during the twentieth century than genes. Genes and their roles in living systems have been typically

[18] In the post-genomic era, researchers have been pushed into adopting a "systemic" perspective, which has given rise to a wave of "systems biology" in the fields of molecular biology, genomics, and proteomics. Systems biology is often presented as a non-reductionistic approach (Chong and Ray 2002; Barabási and Oltvai 2004; Nature 2005). Many genomic researchers seem quite eager, indeed, to declare that they have overcome "fallacies" such as determinism and reductionism (see, e.g., Venter et al. 2001, p. 1348), even though a sort of embarrassed determinism (cf. Leite 2007) lives on in their writings. But it is not clear, at present, what "systems biology" really means in these fields (Keller 2005), and, furthermore, it can be put into question if it is really such a non-reductionistic approach as many of its advocates claim (Bruni 2003; Morange 2006; El-Hani et al. 2009), much in the same sense as systems ecology was previously charged of being nothing but a large-scale reductionistic approach (e.g., Levins and Lewontin 1985; Bergandi 1995).

understood in a rather isolated manner throughout the last century, as shown by the very appeal to the idea that genes could be units of structure, function, and information. After all, this idea is committed to the assumption that genes could have meaning when in isolation. Keller's claim that the twenty-first century will be the century of genetic systems is related to the idea that a reductionist view of genes as having meaning in isolation is to be overcome by a more systemic view. Furthermore, in order to accommodate the challenges posed by biological complexity, she also points in the direction of a process-oriented mode of thinking. She invites us to adopt a new lexicon in which genes are taken to be *verbs*, rather than *substantives*. Genes should be reconceptualized, then, in the context of an understanding of the complex information networks constituting the cell, regarded as a meaning making system that turns nucleotide sequences into genes. In this context, it is important to focus on what *genes can't do* (Moss 2003a), given the hyperbolic view about genes that prevailed in the second half of the twentieth century, as if they were all-powerful controllers of cells and organisms.

From this perspective, genes are conceived in a more dynamic manner and can be even the case that we should say that the cell *genes* (using "gene" as a verb, rather than a substantive), i.e., dynamically combines elements of its memory system—nucleotide sequences—so as to build genes. This may be an apt way of capturing previous ideas, such as Fogle's and Pardini and Guimarães', to the effect that genes are combinations or sets of nucleotide sequences, instead of demarcating genes as entities in DNA. But, perhaps, one could still demarcate *genes* (as substantives) in another component of the cell, such as mature RNA transcripts, as we will mention below.

The proposal of treating genes as processes is further developed by El-Hani et al. (2009), who performed a semiotic analysis of genetic information processing, based on Charles S. Peirce's theory of signs. This analysis led them to the idea that genes can be treated as signs, the genome, as a system of signs, and genetic information, as the action of genes as signs (in Peirce's terms, "semiosis" is used to designate the actions of signs). They offer two interpretations for the results of their semiotic analysis, embracing two distinct ways of conceiving genes.

The major intention of their work is to clarify the meaning of "genetic information", a concept that is used in biology in a largely metaphorical manner and still stands in search of a theory (Griffiths 2001). In order to perform this conceptual elucidation, El-Hani et al. (2009) follow previous indications—such as Oyama's ([1985]2000)—to the effect that we should avoid thinking of genetic information as being stored in DNA sequences.

Since the beginnings of molecular biology, "information" was simply identified with or at least strongly related to sequences of nucleotides in a string of DNA constituting a "gene". Crick (1958, p. 153), for instance, writes that "information means […] the precise determination of sequence, either of bases in the nucleic acid or of amino acid residues in the protein". When information is conceived in terms of 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, through the central dogma, to DNA. When information is conceptualized this way, DNA becomes a sort of reservoir from where all "information" in a cell flows and to which it must be ultimately reduced. We tend to overplay, then, the role of DNA in cell systems, turning it into a complete "program for development" or an all-powerful "controller" of cell metabolism, overlooking that DNA seems to function as a set of data rather than as a program in cell systems (Atlan and Koppel 1990), as a supplier of materials for development rather than as a master agent (or master molecule) in cell

processes (Nijhout 1990), as a bookkeeping entity rather than as a unit of selection in most selective processes in evolution (Gould 2002). Indeed, DNA plays roles that are *obviously important*, but cannot be correctly described by postulating programs or controls realized by this molecule. In short, we should not neglect the fact that *it is not DNA that does things to the cell; rather, it is the cell that does things with DNA* (El-Hani et al. 2009). This is, indeed, one of the major conclusions we can take from developments in the debates around the gene concept in the last three decades, and it can be regarded, accordingly, as an important take-home message for teachers and science educators.[19]

In a first, less radical interpretation, El-Hani et al. (2009) still locate genes in DNA, but as potential signs, whose effects on the cell take place through a semiotic process that irreducibly involves three elements, in the sense of Peirce's interpretation of semiosis: a string of DNA as a sign, a specific sequence of amino acids as an object, and a range of interpretability of signs in DNA, i.e., a set of possibilities of reconstruction of specific amino acid sequences as an interpretant. In this manner, the meaningfulness of genes for cells emerges at a systemic level that is higher than the molecular one, not being reducible to the DNA level. In this interpretation, only genetic information is conceived as a process, coordinated by the cell, taken to be an interpreter of signs found in its memory system, which importantly includes DNA, but not only DNA. This process is not carried out by the cell in an isolated manner, but in the context of its responses to the environment through a set of signaling pathways, which alter patterns of gene expression.[20] Genes are potential signs in the sense that they carry the likelihood that certain effects will take place in their interpreters, by means of a potential semiotic process that can be actualized as information. It is only through the triadic relationship between sign, object, and interpretant that the effects of gene products—functional proteins or RNAs—on the cell can take place.

These authors also offer a more radical interpretation, in which genes are conceived as processes rather than entities in DNA. Genes are conceived as emerging as processes at the level of the systems through which DNA sequences are interpreted, involving both the cellular and the supracellular environment. Thus, genes are not found in DNA itself, but built by the cell at a higher systemic level. In this manner, they walk very close to Griffiths and Neumann-Held's (1999) and Neumann-Held's (2001) proposal of treating genes as processes, and also benefit from Keller's (2005) suggestion of capturing the meaning of "gene" in a verb rather than a substantive, alongside with using the expression "genetic information" to denote the process by means of which the gene has an effect on the cell through a semiotic, irreducibly triadic process involving not only signs in DNA, but also objects and interpretants (defined as above).

Process-oriented views about genes can lead to relevant changes in our way of thinking about genetic phenomena, reinforcing the current trend of shifting from a biological knowledge focused on parts of living systems to systems biology, looking for an understanding of the network of complex interactions between cell components in a processual, non reductionist manner. After all, it is important to appraise how we conceive of systems biology in current times, either as simply the production of a huge catalogue of molecules and interactions, or as a truly systemic understanding of how the biological systems function over and above the molecular level, taking in due account regulation as a top-down effect and the fact that living systems are not merely sacs of molecules, but highly

[19] We added this remark by inspiration of a comment made by a reviewer of the original manuscript, who called attention to this sentence as such a take-home message, stimulating us to highlight it in the paper.

[20] For a semiotic interpretation of signaling pathways, also based on C. S. Peirce's theory of signs, see El-Hani et al. (2007).

organized systems, both temporally and spatially. Models that go far beyond catalogues of molecules and interactions are then required, grasping how the dynamics of the living system as a whole or of modules identifiable in its networks takes place in space and time. The risk that systems biology degenerates into nothing more than a large-scale reductionism results from the very lack of clarity about what "systems biology" really means (Keller 2005). Process-oriented treatments of genes and their roles in living beings can help in avoiding such a reductionist interpretation of systems biology.

4.3 New Languages to Talk About Genetic Systems

At least part of the solution to the gene problem may lie in building a new language to talk about genetic systems. A language embodies a series of operations of distinction in the world of experience. These operations lead to a categorical scheme, in such a manner that to think new languages is to think new ontologies. Languages frame not only our way of speaking about something, but also our way of thinking. We can assume, thus, a close relationship between modes of thinking and ways of speaking, as we see, for instance, in Vygotsky's theory (e.g., Vygotsky 1978). In linguistics, the claim of this relationship has been named the Sapir-Whorf hypothesis, following the works of Benjamin Lee Whorf and Edward Sapir. Whorf expresses it as follows:

> we dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds – and this means largely by the linguistic systems in our minds. We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agreement to organize it in this way – an agreement that holds throughout our speech community and is codified in the patterns of our language (Whorf 1940, pp. 213–14).

Thus, a new language entails new categories to cut up the genetic system (i.e., a new categorical scheme), organizing our understanding into different sets of concepts, and, thus, ascribing new significances to genes, genetic information, and so forth. This may solve, or even dissolve, several problems and limits posed by our current language about genes, impregnated with difficulties that may result, at least in part, from the way we categorize phenomena.

The idea of proposing a new language to speak about genomic elements is not new. Brosius and Gould (1992) already suggested that we needed a new vocabulary to talk about genes, in their case for integrating the studies of evolution and molecular biology. However, we find in Keller and Harel (2007) and Scherrer and Jost (2007a, b) more recent proposals that may have a different fate than Brosius and Gould's, which was entirely ignored, as they themselves anticipated.

4.3.1 Beyond the Gene: The Genetic Functor or Genitor

In 2005, Keller argued for the necessity of building new ways of thinking and speaking in biology, which could give privilege to living systems and their components' interactions, rather than to isolated components. In 2007, in collaboration with David Harel, she proposed a new language to talk about genetic systems. In the construction of this language, they intended to propose an alternative to both the concept and the word "gene", aiming at a proposal that could be expansive, flexible, and rich in expressive power. In this more recent work Keller comes back, thus, to the position found in *The Century of the Gene*, when she proposed that it was time to forge new words to replace the term "gene".

The alternative put forward by Keller and Harel is based on the concept of a genetic functor, or genitor, G, which is defined as a triple $G = (O, D, B)$, being O, an organism of a specified type (i.e., with specified genetic and behavioral properties), D, a dene, and B, a bene. The dene is a statement about O's DNA (or, more precisely, a truth-valued function of O's DNA sequence) and the bene, a statement about O's behavior (more precisely, a truth-valued function of O's temporal life-span).

In this framework, the gene is replaced by a closely related concept, although situated in an entirely different logical level, the dene. Just like the gene, the dene is intended to capture the bases of genetic transmission, but, differently from the gene, it does not denote merely a stretch of DNA. Rather, a dene is a general kind of statement about DNA—what a logician call a predicate or a property—representing much more intricate characteristics of this molecule than those captured by the simple statement that it contains a particular subsequence that is expressed. For each sequence, the dene is—or at least should be—, as a clear and unambiguous statement, either true or false in the sense that the DNA either has the indicated property or not.

The constitutive elements of the dene are found in DNA, but the dene may refer to the entire genome of an organism, or to some part of it, whose boundaries may be fixed or variable. It may also refer to contiguous or disjoint parts of the genome. The dene may contain, moreover, overlapping or iteration of subsequences, and include sequences that possess a functional meaning, even if they are not expressed. It is by means of this flexibility of the dene, as a general statement about DNA, that Keller and Harel intend to accommodate the challenges that put into question the classical molecular concept. The statement a dene makes need not be a simple and direct property of the DNA sequence, but it can rather refer to far more complicated characteristics. It is also important to emphasize that, even though the dene's constitutive elements are found in the DNA, the dene cannot be said to reside in DNA, since it is a statement about this molecule, a designated property of it.

The concept of dene is taken by Keller and Harel as providing a clearer separation—in comparison with the current language about genes—between *structure*, what the organism statically is, what constitutes it, or, in other words, what it materially inherited, and *function*, what the organism dynamically does with what it inherited, i.e., its functionality and behavior. The dene alone says nothing about function, but just makes a statement about DNA as a static entity, a fixed sequence. In this manner, it makes it possible to accommodate the idea of locating agency in living systems at the cellular and supracellular levels, rather than reducing it to the DNA level.

The role of specifying the behavior associated to DNA sequences is ascribed, in Keller and Harel's framework, to the bene, which, together with the dene, form the genitor, G. The genitor expresses the functional relation between D and B. The bene is a statement about how the organism dynamically develops, lives, behaves, etc. We can take as being part of the bene anything that the organism O does that is a manifestation—even if only very indirectly so—of the fact that its DNA possesses the characteristics defined by the dene. In this manner, genetic determinism or other manifestations of gene- or DNA-centric views can be circumvented, since the bene does not simply follow from the dene, but only has the dene as one of its parts, alongside with O's environment, developmental mechanisms, epigenetic processes, and so on. Just like the dene, the bene is a broad and rich statement, referring to complex modal and temporal characteristics of the organism's behavior over time. It cannot be reduced to simple statements, say, about protein synthesis or transcription.

As mentioned above, Keller and Harel relate denes and benes by means of the genitor, or genetic functor, a concept that unifies structure (dene, D) and function (bene, B) in an organism O, $G = (O, D, B)$. In this manner, G connects static and dynamic aspects, providing a more complete understanding of living phenomena.

In their article, Keller and Harel offer a series of examples of the relationship between dene, bene, and genitor. One of them concerns alternative splicing. In this example, the dene refers to any set of mRNA transcripts sewn together to form a protein-coding unit. The dene does not require that the subsequences related to these transcripts be contiguous, and does not forbid their overlapping. Each dene is associated with the corresponding polypeptide.[21] Thus, in a case like alternative splicing, the number of denes far exceeds the number of genes, as currently conceived, coming closer to the size of the metazoan proteome and contributing to a solution of the N value paradox (Claverie 2001, see above). Finally, in alternative splicing, the genitor captures the relation that specifies the components of the dene corresponding to the bene that specifies the production of the polypeptide.

Keller and Harel's proposal of a new language to talk about genetic systems brings a potentially powerful contribution, to the extent that it offers a flexible and rich conceptual structure, which seems capable of accommodating many anomalies challenging our current understanding of genes and, at the same time, avoids gene- or DNA-centric views. The major shortcoming of this approach lies, in turn, in the effort of comprehension that it demands, mostly because the terms introduced—dene, bene, genitor—are quite abstract, and, in fact, use a definition strategy not so commonly employed by the scientific community, namely, that of defining kinds of statements, rather than ontological entities or processes.

Moreover, as the proposed language makes operations of distinction in the world of experience that differ from those found in the categorical structure which we use in the fields of genetics and molecular biology, there is an expected difficulty of intertheoretic translation. A perplexity may be produced in some of us by the concepts of dene, bene, and genitor, due to the hesitations in translating from one categorical scheme to the other. For example, we can face difficulties in translating from this new categorical scheme, with its denes and benes, to the categorical scheme with which we dealt throughout the twentieth century, speaking of genes, phenotype, genotype, among other categories. These are difficulties, however, which typically come along with times of crisis and paradigm change in science, and are far from being surprising. No one can foretell the future of scientific language in the field of genetics, and it may be that, despite eventual difficulties, Keller and Harel's language proves to be so promising that the scientific community engages in the associated change in its way of thinking and speaking about DNA and its role in organisms. There is, however, a competing language in the scenario, which may eventually seem more familiar to geneticists and molecular biologists, to the extent that it shows a closer relationship with the current language used in these fields, as well as with their usual strategies of defining concepts.

To introduce a new way of speaking can indeed solve problems in our understanding, but it is important to keep it as simple as possible. Furthermore, it is a good idea to maintain sufficient bridges between new and older ways of speaking, because more continuity will make it easier to translate from a previous language to a new language, one of

[21] Here we can clearly see how Keller and Harel's dene concept is different from Gerstein and colleagues' treatment of the gene as a union of genomic sequences encoding a coherent set of potentially overlapping functional products. In this latter concept, the gene is taken to be the union of all sequences encoding functional products synthesized by means of the differentially spliced mRNA molecules. The dene, in turn, is a statement about each of the differentially spliced mRNAs.

the most important barriers to a change in our way of speaking that might dissolve or solve problems.

Another difficulty for the acceptance of Keller and Harel's proposal may lie in the fact that they leave no room for the term "gene". Some terms become very ingrained in scientific language, being difficult to displace them, due to the resistance of the scientific community. "Gene" is certainly one of these terms, and, thus, any proposal of change in the language used to speak about genetic systems which does not find a place for it is likely to face opposition. For instance, among a number of book reviews of Keller (2000), where she proposed that the gene concept needed to be cast aside, we found only one which agreed with her proposal (Rios 2004). In all the other reviews and, also, in some scientific papers quoting the book that we examined, the idea was rejected (Coyne 2000; Magurran 2000; Maynard Smith 2000; Hall 2001; Wilkins 2002; Moyle 2002). In the case of the term "gene", it is almost inevitable that one poses the following question: how could we think of genetics research or teaching without genes, given the central role played by the term in this field throughout the twentieth century? As Judson (2001, p. 769) wrote, "… we cannot abandon the term gene and its allies". Symptomatically, we can find in the literature several recent works suggesting that the gene concept continues to be useful in empirical research (e.g., Waters 2004; Weber 2004; Burian 2005), and, even more than that, "… is likely to continue its central role in biology for the foreseeable future" (Knight 2007, p. 303).

4.3.2 Coding Versus Regulation: The Gene, the Genon, and the Transgenon

Scherrer and Jost (2007a, b) take as a starting point the thesis that there are two distinct aspects involved in the production of polypeptides—coding and regulation—to arrive at the proposal that we need distinct terms to account for them. From this perspective, the problem resides in ascribing to a single term, "gene", a very large diversity of meanings. To advance in solving the gene problem, thus, we should perform conceptual analysis, and, consequently, new operations of distinction, which can be grasped, in turn, in the construction of new terms.

They propose that our understanding of gene expression should be broken up into gene function, on the one hand, and the mechanisms of storage and expression, on the other. The gene is mostly related to the functional aspect. They elaborate, then, on the thesis that we cannot found at the DNA level a unit that can account for gene function, and propose, rather, that the gene emerges at the level of mature RNAs. Here, we have a revolutionary claim about genes, demanding that putative advantages and difficulties of locating genes in RNA, rather than in DNA, be discussed. Nevertheless, we will develop this discussion elsewhere, since it would take us away from the main goal of this paper.

If we focus on protein-coding genes, they emerge when a processed, uninterrupted sequence of nucleotides appears, which can act as a structural unit for translation. According to this view, the uninterrupted mRNA sequence that results from processing is the unit of function and genetic analysis, since it constitutes, when reliably translated, the equivalent of the polypeptide chain that will be produced.

If we now consider RNA genes, the situation is a bit more complex, since we might locate the gene at the DNA level when no processing is involved in the synthesis of a functional RNA. However, if there is processing, as it is the case of most functional RNAs, the gene will be also located at the RNA level. In fact, as is the case for transcripts that are destined to be translated into proteins, those that will exercise functions in the form of RNA molecules commonly experience post-transcriptional editing. For example, tRNAs

(e.g., Abelson et al. 1998), rRNAs (e.g., Cech et al. 1981), siRNAs,[22] miRNAs[23] (e.g., Pontes and Pikaard 2008), and snoRNAs[24] (e.g., Giorgi et al. 2001) all have been documented to experience either splicing or editing events, subsequent to the generation of the primary transcript. This implies that the arguments presented above are also relevant to transcripts that will not be translated into proteins, and, for the sake of coherence and generality, it is better to locate the RNA gene at the functional RNA level, claiming that, in the exceptional cases where no processing is involved, what we will see is equivalence between genomic domain[25] (in DNA) and gene (in RNA). We can, thus, simply state that the gene is the functional RNA, when we are dealing with transcripts that will not be translated into protein. But it is important to remember that, in this case, no coding is really taking place, since the synthesis of functional RNAs do not depend on any set of rules to translate from one kind of polymer to another, such as the rules of the genetic code.

In sum, the uninterrupted nucleic acid stretch of the coding sequence, in protein-coding genes, or of the functional ribonucleotide sequence, in the case of RNA genes, should be, for Scherrer and Jost, "the unique and exclusive definition and meaning of the term 'gene'" (Scherrer and Jost 2007a, p. 3) They propose the following definition for "gene", putting much more emphasis on protein-coding than RNA genes (something that should be avoided, in our view):

> … the uninterrupted nucleic acid stretch of the coding sequence in the mRNA that corresponds to a polypeptide or another functional product; thus, in eukaryotes typically not yet present at DNA level, but assembled from gene fragments (exons) in course of RNA processing (Scherrer and Jost 2007b, p. 106).

The second aspect, related to the regulation of gene expression, concerns the process of creation of the gene in RNA out of pieces in the genome (or, more precisely, in genomic domains in DNA). When Scherrer and Jost focus on this process, they stress the role of a programme that comes along with the transcript and "… secures the generation of the gene, in the cellular space and in time, through the many steps of gene expression" (2007a, p. 3).[26] This programme is conceived by them as the "genon" (a contraction of the terms

[22] Small interfering RNA: a class of small double-stranded RNA molecules, which play several roles in the cell, including a key involvement in RNA interference, an important phenomenon of genetic regulation discovered in the 1990s. RNA interference silences gene expression in a highly specific manner.

[23] Micro-RNAs: short RNA molecules that act as post-transcriptional regulators by binding to complementary sequences on target mRNAs, usually resulting in repression of protein synthesis and, thus, gene silencing. Although the first miRNAs were characterized in the early 1990s, their recognition as a distinct class of biologic regulators only took place in the early 2000s. Nowadays, they were shown to play several functions in gene repression and activation.

[24] Small nucleolar RNAs: a class of small RNA molecules that guide chemical modifications of other RNAs, such as rRNAs and tRNAs.

[25] A "genomic domain" is defined by Scherrer and Jost (2007b, p. 106) as a "DNA domain containing fragments of one or several genes coordinated by *cis* controls, […] often unit of transcription and, in some cases, of replication".

[26] The notion of 'program', particularly when conceived in terms of "genetic programs", is highly controversial (e.g., Oyama [1985]2000; Nijhout 1990; Moss 1992; Griffiths and Neumann-Held 1999; Keller 2000), but we will not pursue this discussion here, since it would take us away from our major goals in this paper. Scherrer and Jost do not elaborate on the concept of "programme". They remark, however, that the genon and transgenon constitute a flexible, not rigidly defined program, to the extent that epigenetic mechanisms of gene expression and transmission modify both the genon and its precursors at the DNA level. In these modifications, the genon and transgenon may be modified with no changes being made at the DNA level. The genon, for instance, can be changed by epigenetic modifications such as DNA methylation, while the transgenon can be modified by the addition or elimination of factors originated either in the genome or in the environment, according to cell compartment, physiological context, cell age, etc.

"gene" and "operon"). The genon amounts to the additional information needed to gene expression and is contained within each RNA as an ensemble of signals, potential binding sites for regulatory proteins, RNAs or ribonucleoproteins, both added to and superimposed onto the coding sequence. The genon is unique to each distinct mRNA and polypeptide, and, thus, to each gene, although the same DNA segments can be used in different combinations along the expression pathways of similar or different genes.[27] When the cell builds a gene out of DNA sequences, it also builds a corresponding genon.

The genon is found in the same chromosome from which the pre-RNA is transcribed, i.e., it is a *cis* acting regulatory programme.[28] There is, however, also an ensemble of regulatory factors codified in other chromosomes that potentially recognize and act on the signals in *cis* found in a given genon. This ensemble is called the "transgenon".[29] As the genon of an mRNA is immersed into a pool of trans-acting factors (holo-transgenon) which may be capable of recognizing signals or "oligomotifs" in the nucleotide sequence, and, thus, may act on genons and genes, a specific transgenon is selected out by each genon. It is from this genon-transgenon specific interaction that the regulation of gene expression emerges.

The gene is constructed in RNA processing due to a variety of processes, including the alternative splicing of several exons, directed by the genon, and mRNA editing, directed by the transgenon, if we consider protein-coding genes. As we said above, the gene emerges at the mature RNA level, and, some time after translation, loses its function and is degraded alongside with the genon. If we consider the sequences that come to constitute the gene and genon at the RNA, in most cases they only exist in DNA as fragments. Genes are typically created by the cell from its parts encoded in genomic domains in DNA. As the gene is not located at the DNA level, but only at the mature RNA level, most of the anomalies faced by the classical molecular concept are overcome by Scherrer and Jost's proposal, with some remaining exceptions, such as alternative modes of translation.

The proposal developed by Scherrer and Jost (2007a, b) substantially expands the vocabulary related to genes. As Pearson (2006) remarks, this seems to be a strategy used by many scientists in the face of the challenges to the gene concept, since it has become usual to add adjectives to the term "gene" in order to overcome the limits of a single meaning ascribed to a term that needs to refer to a widely diversified range of entities and processes. We should reflect, however, about the impact of introducing a richer vocabulary to account for what was expressed by a single term. On the one hand, it may be difficult to convince the scientific community to adopt the series of new terms in a novel language to speak about genetic systems. On the other hand, Scherrer and Jost's framework has, in our view, clear advantages in relation to Keller and Harel's, since they preserve the term "gene", introduce other terms that are not as hard to understand as those offered by the latter, and their vocabulary keeps more contact with current genetic language and strategies for defining terms.

Anyway, to overcome the problems challenging the gene concept indeed requires efforts of conceptual analysis and change, aiming at distinguishing between the varied meanings that came to be amalgamated in the term "gene" throughout the second half of the twentieth century. And, when we perform conceptual analysis, it is just natural that we increase the vocabulary of a science. After all, new operations of distinction demand new operations of naming, in order to add clarity to the use of the renewed conceptual structure

[27] Evidently, this is highly reminiscent of Fogle's and Pardini and Guimarães' accounts about the gene.

[28] The prefix 'cis' is used to denote a sequence that is in the same DNA molecule, in the same chromosome, in relation to another sequence of interest.

[29] The prefix 'trans' is used to denote a sequence (or its product) at a different DNA molecule or chromosome, in relation to a sequence of interest.

of a scientific field, making it possible to communicate by means of the new terms, avoiding semantic confusions and troubling ambiguities. In sum, to change our ways of speaking about genes may be an inevitable price to pay in order to overcome the crisis of this concept. Now, it is just a matter of sailing the waves of history to see which new language and framework to think about genes, if any, will eventually be accepted by the scientific community.

5 Concluding Remarks

The significance of the debates about the gene concept to science education becomes clear when we consider that both high school textbooks (Gericke and Hagberg 2007a, b, 2010a, b; Santos and El-Hani 2009; Santos et al. in press) and higher education textbooks from relevant fields, such as cell and molecular biology (Pitombo et al. 2008) and introductory biology (Flodin 2009), still represent genes—not only, but to a great extent—in accordance with the classical molecular concept. Moreover, they typically do not engage with the challenges to our usual way of understanding genes, even when they discuss the phenomena related to them—such as alternative splicing and overlapping genes. Moreover, studies about Brazilian college students' views about genes suggest that we may have reasons to infer that what is found in textbook analysis is also reflected on how the topic is being treated in classrooms. In two different studies, ideas related to the classical molecular concept prevailed among biological and medical sciences students (Joaquim 2009; Meyer 2010). Moreover, genetics teaching in two different universities did not have much consequence regarding changes in the prevalence of the commitment to the classical molecular gene or the informational conception (Joaquim 2009), and, even after explicit teaching about the crisis of the gene concept and the proposals to overcome it, the classical molecular concept was still accepted by many students 2 months after the intervention (Meyer 2010). Another troubling finding of these studies is that genetic deterministic views are highly represented among both textbooks and students, and can be related, at least partly, to the prevalence of a conflation between gene-P, as an instrumental gene concept used in school science mostly in the context of classical genetics and in association with pedigree analysis, and gene-D, as a more realist concept mostly related in school science to molecular-based treatments of genetic systems.

Finally, a number of studies have shown that distinct ideas about genes, related to models built in different historical contexts, in order to deal with distinct research questions, and not entirely compatible with each other, are indiscriminately mixed up in textbooks, and, also, in the views reported by students (Gericke and Hagberg 2007a, b, 2010a, b; Pitombo et al. 2008; Santos and El-Hani 2009; Santos et al. in press; Joaquim 2009; Meyer 2010). This conflation leads not only to consequential problems in students' understanding of genes and their role in living beings—such as the commitment to a hyperbolic, overextended view of what DNA and genes do in cell systems—, but also has implications to gene talk in society, reinforcing its strong commitment to genetic determinism. Even though hybridization between ideas originating from different historical models about genes and gene function takes place as part of the very construction of scientific knowledge, in biological education it often becomes a source of problems for students' learning, due to the frequent absence of a proper historical and philosophical background in teaching about those models and putative relationships between them.

All these findings suggest that we need to reconsider the way we have been teaching about genes, attempting to give more attention to the current situation of the classical

molecular concept, instead of just teaching about it as if it was as accepted and coherent as it was in the past. At least, the fact that there are serious debates about what is a gene in the scientific community deserves attention in genetics teaching, even at the high school level. That is, the gene concept should be currently treated as a controversial subject matter in the structure of biological thought.

Needless to say, we should consider the effective possibility of transposing to school science the current understanding of the anomalies challenging the classical molecular concept and the alternatives to this way of understanding genes. In the case of high school biological education, we think it is possible to create conditions for the students to understand that, even though the classical molecular concept has been quite important in the history of biology, it has ended up showing consequential limitations. That is, the recognition of the crisis of the gene concept, more specifically, of the classical molecular concept, seems to be a reasonable teaching and learning goal at the high school level. Some anomalies can be discussed in some detail at this level, such as overlapping genes and alternative splicing, while others cannot, such as, say, mRNA editing or genomic rearrangements. Moreover, the concepts of gene-P and gene-D, as well as their clear demarcation, alongside with a critique of genetic determinism, find adequate places in a high school genetics curriculum.

In higher education, including teacher education, a more thorough consideration of the crisis of the gene concept can be achieved, with wider and deeper discussions of the anomalies challenging the classical molecular concept, and even an examination of some alternatives to this concept. Obviously, the depth of this discussion will depend on whether we are teaching an introductory biology course or a more specific discipline, such as genetics or cell and molecular biology, or whether we are teaching to biology majors or not, or whether we are involved in a teacher education program or not. Be that as it may, the treatment of the gene problem and the current state of affairs, with a proliferation of proposals to solve it, can be certainly taken to a deeper level than at high school.

Another relevant requirement for a proper treatment of genes in school science at both high school and higher education lies in the need of clearly demarcating different models about genes and their function, so as to make it clear what are their domains of applicability, and, thus, the limits beyond which they are no longer appropriate. The importance of this requirement is reinforced by the proposal that we need a plurality of views about genes, rather than a single, all-encompassing gene concept (Burian 2004; El-Hani 2007). Such conceptual variation cannot mean, however, that "it may not be important to know what the precise meaning of 'gene' is" (Knight 2007, p. 300). To entertain the importance of a clear treatment of different gene concepts and models, and their boundaries, we need just to rephrase this statement by considering a plurality of ways of understanding genes: even though it is not really important to provide a single precise meaning of "gene", we need, still, to provide a clear and precise understanding of the several different meanings of "gene", since they cannot be all put to each and every use.

A clearer treatment of different historical models of genes and their function would help putting into question, for instance, genetic deterministic views, as they often follow, as discussed above, from indiscriminate combination of features related to different models. Moreover, this treatment would make it possible to address Nature of Science (NOS) contents in connection with the history of the gene concept. After all, the transition from the understanding of genes in classical genetics to the molecular gene with the advent of molecular biology, and then to the crisis of the gene concept and the various approaches proposed over the last years is a very interesting case of conceptual change, as well as of how science studies and represents unobservable entities.

We should say, however, that at this stage of our research on the treatment of genes in school science, all these ideas are just educated guesses, which can be used as design principles for teaching interventions, but have not been empirically tested in the classroom. Our intention in this paper was to provide an informative overview and discussion of recent views about genes, probably not very familiar to many science education researchers and teachers, since we think they can bring relevant contributions to genetics teaching, in particular, to a more critical treatment of genes and their role in living systems, either at higher education (including teacher education) or at the high school level. This is a subject matter in which the contributions of philosophy of biology to science teaching are particularly illuminating, as we expect to have shown in the paper. When it comes to genes, to give due attention to the historical and philosophical aspects related to this central concept in biology is a key requisite for successful teaching and learning about them.

In Table 1, we summarize the views about genes reviewed and discussed in this paper, indicating to which school level we consider they can be adequately transposed.

Table 1 Views about genes and adequate schooling level for their transposition

Views about genes	Explanation	Adequate schooling level[a]
Classical molecular gene concept (Griffiths and Neumann-Held 1999; Stotz et al. 2004)	A stretch of DNA that encodes a functional product, a single polypeptide chain or RNA molecule Genes as structural, functional, and informational units	High school Higher education Teacher education
Crisis of the gene concept—Anomalies challenging the classical molecular gene concept (Pardini and Guimarães 1992; El-Hani 2007)	Three kinds of anomalies and their consequences for a unitary relationship between genes, gene products, and gene function: (i) one-to-many correspondences between DNA segments and RNAs/polypeptides (e.g., alternative splicing) (ii) many-to-one correspondences between DNA segments and RNAs/polypeptides (e.g., genomic rearrangements) (iii) lack of correspondence between DNA segments and RNAs/polypeptides (e.g., mRNA editing)	High school—partially, for instance, alternative splicing Higher education Teacher education
Process molecular genes (Griffiths and Neumann-Held 1999; Neumann-Held 2001)	Gene as the recurring process that leads to the temporally and spatially regulated expression of a particular polypeptide product	Higher education Teacher education
Gene-P (Moss 2001, 2003a)	Gene as determinant of phenotypes or phenotypic differences Statements of the form "*X* is a gene for *Y*"; *X*, particular gene; *Y*, human disorder or trait Instrumental concept, not accompanied by hypothesis of correspondence to reality Not to be conflated with gene-D	High school Higher education Teacher education
Gene-D (Moss 2001, 2003a)	Gene as developmental resource, real entity defined by molecular sequence in DNA Developmental resource for the construction of traits, along with equally important resources (epigenetic and environmental factors) It does not determine phenotypic trait Not to be conflated with gene-P	High school Higher education Teacher education

Table 1 continued

Views about genes	Explanation	Adequate schooling level[a]
Genes as sets of domains in DNA (Fogle 1990, 2000)	Gene constructed from assemblage of embedded, tandem, and overlapping domains in DNA Domains are nucleotide sequences distinguished by structural properties and/or activities: exons, introns, promoters, enhancers, operators, etc.	Higher education Teacher education
Systemic gene concept (Pardini and Guimarães 1992)	Gene as combination of nucleic acid (DNA or RNA) sequences, defined by the cellular system and corresponding to a product (RNA or polypeptide)	Higher education Teacher education
ENCODE definition of the gene (Gerstein et al. 2007)	Gene as union of genomic sequences encoding a coherent set of potentially overlapping functional products	Higher education Teacher education
Semiotic analysis of genes and genetic information (El-Hani et al. 2009)	Genes as potential signs, actualized by semiotic process, the action of genes as signs, which is genetic information More conservative interpretation, genes still in DNA, as potential signs Less conservative interpretation, genes also treated as processes, just as genetic information	Higher education
Genitor (genetic functor), Dene, and Bene (Keller and Harel 2007)	Genitor (G) = (O, D, B); O, organism of a specified type; D, dene; B, bene Dene is statement about O's DNA; it does not say, by itself, anything about function, it is just a statement about DNA as static entity, fixed sequence Bene is statement about O's behavior, statement about how the organism dynamically develops, lives, behaves, etc. The genitor unifies structure (dene, D), as static aspect, and function (bene, B), as dynamic aspect, in organism O	Higher education
Gene, genon, and transgenon (Scherrer and Jost 2007a, b)	Gene as the uninterrupted nucleic acid stretch of the coding sequence in mature mRNA, in protein-coding genes, or of the functional ribonucleotide sequence, in RNA genes (rRNAs, tRNAs, miRNAs, etc.) Genon as the *cis* acting programme that regulates gene transcription Transgenon as the ensemble of regulatory factors codified in other chromosomes that potentially recognize and act on the signals in *cis* found in a genon The mature mRNA has both a genon and a gene. A specific transgenon is selected out by each genon, and the genon-transgenon specific interaction regulates gene expression	Higher education Teacher education

[a] Although teacher education is typically a part of higher education, we treat them separately in the table in order to make it clear that some views about genes are more adequate to be transposed to the education of biologists, physicians, and other health sciences students. Accordingly, when we refer to "higher education", we are focusing on biology and health sciences courses. When we refer to "teacher education", we are concerned with biology teacher education. It follows that we are considering that in primary school more basic concepts related to genetics than a discussion of what genes are should play a major role

Acknowledgments We would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES) and the State of Bahia Foundation for the Support of Research (FAPESB) for graduate studies grants, and the National Council for Scientific and Technological Development (CNPq) and FAPESB for financial support. We also thank CNPq for a grant for productivity in research. We are also indebted to Vanessa Carvalho dos Santos and Leyla Mariane Joaquim for suggestions about the manuscript.

References

Abelson, J., Trotta, C. R., & Li, H. (1998). tRNA splicing. *The Journal of Biological Chemistry, 273*, 12685–12688.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics, 25*, 25–29.

Ast, G. (2004). How did alternative splicing evolve? *Nature Reviews Genetics, 5*, 773–782.

Atlan, H., & Koppel, M. (1990). The cellular computer DNA: Program or data? *Bulletin of Mathematical Biology, 52*, 335–348.

Barabási, A. L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics, 5*, 101–113.

Bergandi, D. (1995). "Reductionist Holism": An oxymoron or a philosophical chimera of E. P. Odum's systems ecology? *Ludus Vitalis: Journal of Philosophy of Life Sciences, III*, 145–180.

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

Brosius, J., & Gould, S. J. (1992). On "Genomenclature": A comprehensive (and respectful) taxonomy for pseudogenes and other "Junk DNA". *Proceedings of the National Academy of Sciences of the United States, 89*, 10706–10710.

Bruni, L. E. (2003). *A sign-theoretic approach to biotechnology*. Copenhagen: Institute of Molecular Biology, University of Copenhagen. Ph.D. Thesis.

Burian, R. M. (1985). On conceptual change in biology: The case of the gene. In D. J. Depew & B. H. Weber (Eds.), *Evolution at a crossroads: The new biology and the new philosophy of science* (pp. 21–24). Cambridge, MA: The MIT Press.

Burian, R. M. (2004). Molecular epigenesis, molecular pleiotropy, and molecular gene definitions. *History and Philosophy of Life Sciences, 26*, 59–80.

Burian, R. M. (2005). *The epistemology of development, evolution and genetics*. Cambridge: Cambridge University Press.

Carthew, R. W. (2006). Gene regulation by MicroRNAs. *Current Opinion in Genetics and Development, 16*, 203–208.

Cech, T. R., Zaug, A. J., & Grabowski, P. J. (1981). In vitro splicing of the ribosomal RNA precursor of *Tetrahymena*: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell, 27*, 487–496.

Celotto, A., & Graveley, B. (2001). Alternative splicing of the *Drosophila* DSCAM pre-mRNA is both temporally and spatially regulated. *Genetics, 159*, 599–608.

Chien, K. R. (2007). MicroRNAs and the tell-tale heart. *Nature, 447*, 389–390.

Chong, L., & Ray, L. B. (2002). Whole-istic biology. *Science, 295*, 1661.

Claverie, J.-M. (2001). What if there are only 30,000 human genes? *Science, 291*, 1255–1257.

Cooper, M. D., & Alder, M. N. (2006). The evolution of adaptive immune systems. *Cell, 124*, 815–822.

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

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

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

El-Hani, C. N., Arnellos, A., & Queiroz, J. (2007). Modeling a semiotic process in the immune system: Signal transduction in B-cell activation. *TripleC-Cognition, Communication, Co-operation, 5*, 24–36.

El-Hani, C. N., Queiroz, J., & Emmeche, C. (2009). *Genes, information, and semiosis*. Tartu: Tartu University Press (Tartu Semiotics Library).

Epp, C. D. (1997). Definition of a gene. *Nature, 389*, 537.

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

Falk, R. (2000). The gene—A concept in tension. In P. Beurton, R. Falk, & H.-J. Rheinberger (Eds.), *The concept of the gene in development and evolution* (pp. 317–348). Cambridge: Cambridge University Press.

Falk, R. (2001). Can the norm of reaction save the gene concept? In R. S. Singh, C. B. Krimbas, D. B. Paul, & J. Beatty (Eds.), *Thinking about evolution: Historical, philosophical and political perspectives* (pp. 119–140). New York, NY: Cambridge University Press.

Flodin, V. (2009). The necessity of making visible concepts with multiple meanings in science education: The use of the gene concept in a biology textbook. *Science & Education, 18*, 73–94.
Fogle, T. (1990). Are genes units of inheritance? *Biology and Philosophy, 5*, 349–371.
Fogle, T. (2000). The dissolution of protein coding genes. In P. Beurton, R. Falk, & H.-J. Rheinberger (Eds.), *The concept of the gene in development and evolution* (pp. 3–25). Cambridge: Cambridge University Press.
Gelbart, W. (1998). Databases in genomic research. *Science, 282*, 659–661.
Gericke, N. M., & Hagberg, M. (2007a). Definition of historical models of gene function and their relation to students' understandings of genetics. *Science & Education, 16*, 849–881.
Gericke, N. M., & Hagberg, M. (2007b). The phenomenon of gene function in textbooks for upper secondary school in Sweden—A comparative analysis with historical models of gene function. In *Proceedings of the IOSTE International meeting on critical analysis of school science textbooks* (pp. 554–563). Hammamet, Tunisia: University of Tunis.
Gericke, N. M., & Hagberg, M. (2010a). Conceptual incoherence as a result of the use of multiple historical models in school textbooks. *Research in Science Education, 40*, 605–623.
Gericke, N. M., & Hagberg, M. (2010b). Conceptual variation in the depiction of gene function in upper secondary school textbooks. *Science & Education, 19*, 963–994.
Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., et al. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Research, 17*, 669–681.
Giorgi, C., Fatica, A., Nagel, R., & Bozzoni, I. (2001). Release of U18 snoRNA from its host intron requires interaction of Nop1p with the Rnt1p endonuclease. *The EMBO Journal, 20*, 6856–6865.
Gould, S. J. (2002). *The structure of evolutionary theory*. Cambridge, MA: Harvard University Press.
Graveley, B. R. (2001). Alternative splicing: Increasing diversity in the proteomic world. *Trends in Genetics, 17*, 100–107.
Gray, R. D. (1992). Death of the gene: Developmental systems fight back. In P. E. Griffiths (Ed.), *Trees of life: Essays in the philosophy of biology* (pp. 165–209). Dordrecht: Kluwer.
Griffiths, P. E., & Neumann-Held, E. (1999). The many faces of the gene. *BioScience, 49*, 656–662.
Griffiths, P. E. (2001). Genetic information: A metaphor in search of a theory. *Philosophy of Science, 68*, 394–403.
Guimarães, R. C., & Moreira, C. H. C. (2000). O Conceito Sistêmico de Gene—Uma Década Depois. In I. M. L. D'Ottaviano & I. C. Q. Gonzáles (Eds.), *Auto-organização: Estudos Interdisciplinares* (pp. 249–280). Campinas: UNICAMP.
Hall, B. K. (2001). The gene is not dead, merely orphaned and seeking a home. *Evolution and Development, 3*, 225–228.
Hamer, D. H., & Copeland, P. (1994). *The science of desire: The search for the gay gene and the biology of behavior*. New York, NY: Simon & Schuster.
Hamer, D. H., Hu, S., Magnuson, V. L., Hu, N., & Pattatucci, A. M. L. (1993). A linkage between DNA markers on the X chromosome and male sexual orientation. *Science, 261*, 321–327.
Hanson, M. R. (1996). Protein products of incompletely edited transcripts are detected in plant mitochondria. *The Plant Cell, 8*(1), 1–3.
Hendrickson, D. G., Hogan, D. J., McCullough, H. L., Myers, J. W., Herschlag, D., Ferrell, J. E., et al. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human MicroRNA. *PLoS Biology, 7*, e1000238.
Hilgers, V., Bushati, N., & Cohen, S. M. (2010). Drosophila MicroRNAs 263a/b confer robustness during development by protecting nascent sense organs from apoptosis. *PLoS Biology, 8*, e1000396.
Ideker, T., Galitski, T., & Hood, L. (2001). A new approach to decoding life: Systems biology. *Annual Review of Genomics and Human Genetics, 2*, 343–372.
International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*, 860–921.
Joaquim, L. M. (2009). *Gene: Questões Epistemológicas, Conceitos Relacionados e Visões de Estudantes de Graduação*. Salvador: Graduate Studies Program in History, Philosophy, and Science Teaching, Federal University of Bahia and State University of Feira de Santana. Master's thesis.
Judson, H. F. (2001). Talking about the genome. *Nature, 409*, 769.
Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., et al. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Research, 14*, 331–342.
Karres, J. S., Hilgers, V., Carrera, I., Treisman, J., & Cohen, S. M. (2007). The conserved MicroRNA MiR-8 tunes atrophin levels to prevent neurodegeneration in *Drosophila*. *Cell, 131*, 136–145.
Kay, L. E. (2000). *Who wrote the book of life? A history of the genetic code*. Stanford, CA: Stanford University Press.

Keller, E. F. (2000). *The century of the gene*. Cambridge, MA: Harvard University Press.
Keller, E. F. (2005). The century beyond the gene. *Journal of Biosciences, 30*, 3–10.
Keller, E. F., & Harel, D. (2007). Beyond the gene. *PLoS One, 2*, e1231.
Kendler, K. S. (2005). "A Gene for": The nature of gene action in psychiatric disorders. *American Journal of Psychiatry, 162*, 1243–1252.
Khachane, A. N., & Harrison, P. M. (2009). Assessing the genomic evidence for conserved transcribed pseudogenes under selection. *BMC Genomics, 10*, 435.
Kitcher, P. (1982). Genes. *British Journal for the Philosophy of Science, 33*, 337–359.
Knight, R. (2007). Reports of the death of the gene are greatly exaggerated. *Biology and Philosophy, 22*, 293–306.
Leite, M. (2007). *Promessas do Genoma*. São Paulo: UNESP.
Levins, R., & Lewontin, R. C. (1985). *The dialectical biologist*. Cambridge, MA: Harvard University Press.
Lev-Maor, G., Sorek, R., Levanon, E. Y., Paz, N., Eisenberg, E., & Ast, G. (2007). RNA-editing-mediated exon evolution. *Genome Biology, 8*, R29.
Lewontin, R. C. (1983). The organism as the subject and object of evolution. *Scientia, 118*, 63–83.
Lewontin, R. C. (2000). *The triple helix: Gene, organism, and environment*. Cambridge, MA: Harvard University Press.
Li, M., Wang, I. X., Li, Y., Bruzel, A., Richards, A. L., Toung, J. M., et al. (2011). Widespread RNA and DNA sequence differences in the human transcriptome. *Science, 333*, 53–58.
Magen, A., & Ast, G. (2005). The importance of being divisible by three in alternative splicing. *Nucleic Acids Research, 33*, 5574–5582.
Magurran, A. (2000). Backseat drivers, review of the century of the gene by E. F. Keller. *New York Times Book Reviews*, 10 December, p. 26.
Maynard Smith, J. (2000). The Cheshire cat's DNA. *The New York Review of Books, 47*, 43–46.
Meyer, L. M. N. (2010). Como Ensinar a Estudantes Universitários de Ciências Biológicas e Ciências da Saúde sobre a Crise do Conceito de Gene? Salvador: Graduate Studies Program in History, Philosophy, and Science Teaching, Federal University of Bahia and State University of Feira de Santana. Master's thesis.
Morange, M. (2006). Post-genomics, between reduction and emergence. *Synthese, 151*, 355–360.
Moss, L. (1992). A kernel of truth? On the reality of a genetic program. *PSA, 1*, 335–348.
Moss, L. (2001). Deconstructing the gene and reconstructing molecular developmental systems. In S. Oyama, P. E. Griffiths, & R. D. Gray (Eds.), *Cycles of contingency: Developmental systems and evolution* (pp. 85–97). Cambridge, MA: MIT Press.
Moss, L. (2003a). *What genes can't do*. Cambridge, MA: The MIT Press.
Moss, L. (2003b). One, two (too?), many genes? *Quarterly Review of Biology, 78*, 57–67.
Moyle, L. (2002). Most ingenious: Troubles and triumphes of a century of genes. *Biology and Philosophy, 17*, 715–727.
Murre, C. (2007). Epigenetics of antigen-receptor gene assembly. *Current Opinion in Genetics and Development, 17*, 415–421.
Nature. (2005). In pursuit of systems (editorial). *Nature, 435*, 1.
Neumann-Held, E. (1999). The gene is dead—Long live the gene: Conceptualizing genes the constructionist way. In P. Koslowski (Ed.), *Sociobiology and bioeconomics: The theory of evolution in biological and economic thinking* (pp. 105–137). Berlin: Springer.
Neumann-Held, E. (2001). Let's talk about genes: The process molecular gene concept and its context. In S. Oyama, P. E. Griffiths, & R. D. Gray (Eds.), *Cycles of contingency: Developmental systems and evolution* (pp. 69–84). Cambridge, MA: MIT Press.
Nijhout, H. F. (1990). Metaphors and the role of genes in development. *BioEssays, 12*, 441–446.
Niwa, R., & Slack, F. J. (2007). The evolution of animal MicroRNA function. *Current Opinion in Genetics and Development, 17*, 145–150.
Oyama, S. ([1985]2000). *The ontogeny of information: Developmental systems and evolution* (2nd ed.). Cambridge: Cambridge University Press.
Oyama, S., Griffiths, P. E., & Gray, R. D. (Eds.). (2001). *Cycles of contingency: Developmental systems and evolution*. Cambridge, MA: The MIT Press.
Pardini, M. I. M. C., & Guimarães, R. C. (1992). A systemic concept of the gene. *Genetics and Molecular Biology, 15*, 713–721.
Pearson, H. (2006). What is a gene? *Nature, 441*, 399–401.
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*, 219–234.
Pontes, O., & Pikaard, C. S. (2008). siRNA and miRNA processing: New functions for cajal bodies. *Current Opinion in Genetics and Development, 18*, 197–203.

Portin, P. (1993). The concept of the gene: Short history and present status. *Quarterly Review of Biology, 56*, 173–223.

Rheinberger, H.-J. (2000). Gene concepts: Fragments from the perspective of molecular biology. In P. Beurton, R. Falk, & H.-J. Rheinberger (Eds.), *The concept of the gene in development and evolution* (pp. 219–239). Cambridge: Cambridge University Press.

Riordan, J., Rommens, J., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., et al. (1989). Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA. *Science, 245*, 1066–1073.

Rios, R. I. (2004). O Início do Fim do Gene, Resenha de Keller, E. F. O Século do Gene. *Ciência Hoje, 34*, 72–73.

Santos, V. C., & El-Hani, C. N. (2009). Idéias sobre Genes em Livros Didáticos de Biologia do Ensino Médio publicados no Brasil. *Revista Brasileira de Pesquisa em Educação em Ciências, 9*(1), a6.

Santos, V. C., Joaquim, L. M., & El-Hani, C. N. (in press). Hybrid Deterministic views about genes in biology textbooks: A key problem in genetics teaching. *Science & Education* doi:10.1007/s11191-011-9348-1

Scherrer, K., & Jost, J. (2007a). The gene and the genon concept: A functional and information-theoretic analysis. *Molecular System Biology, 3*, 1–11.

Scherrer, K., & Jost, J. (2007b). The gene and the genon concept: Coding versus regulation. A conceptual and information-theoretic analysis storage and expression in the light of modern molecular biology. *Theory in Biosciences, 126*, 65–113.

Smith, M., & Adkison, L. R. (2010). Updating the model definition of the gene in the modern genomic era with implications for instruction. *Science & Education, 19*, 1–20.

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

The ENCODE Project Consortium. (2007). The ENCODE (ENCyclopedia of DNA elements) project. *Science, 306*, 636–640.

Turner, W. J. (1995). Homosexuality, type 1: An Xq28 phenomenon. *Archives of Sexual Behavior, 24*, 109–134.

Uney, J. B., & Lightman, S. J. (2006). MicroRNAs and osmotic regulation. *Proceedings of the National Academy of Sciences of the United States, 103*, 15278–15279.

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J. J., & Lötvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature Cell Biology, 9*, 654–659.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science, 291*, 1305–1351.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological process*. Cambridge, MA: Harvard University Press.

Wang, W., Zhang, J., Alvarez, C., Llopart, A., & Long, M. (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*, 1294–1301.

Waters, C. K. (1994). Genes made molecular. *Philosophy of Science, 61*, 163–185.

Waters, C. K. (2004). What was classical genetics? *Studies in the History and Philosophy of Science C, 35*, 783–809.

Weber, M. (2004). *Philosophy of experimental biology*. Cambridge: Cambridge University Press.

Whorf, B. L. (1940). Science and linguistics. *Technology Review, 42*(229–231), 247–248.

Wilkins, A. S. (2002). Grappling with developmental complexity. *BioEssays, 24*, 1193–1195.

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.

Wu, L., Fan, J., & Belasco, J. G. (2006). MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States, 103*, 4034–4039.

Zhang, B., Pan, X., Cobb, G. P., & Anderson, T. A. (2007). MicroRNAs as oncogenes and tumor suppressors. *Developmental Biology, 302*, 1–12.