Science Education International
Vol. 19, No. 2, June 2008, pp. 219-234

International Council of Associations in Science Education

Gene Concepts in Higher Education Cell and Molecular Biology Textbooks

PITOMBO MAIANA ALBUQUERQUE (maianapit@gmail.com), ANA MARIA ROCHA de ALMEIDA (anagaivota@gmail.com), and NIÑO CHARBEL EL-HANI (charbel.elhani@pesquisador.cnpq.br), Institute of Biology, Federal University of Bahia, Brazil.

ABSTRACT: Despite being a landmark of 20th century biology, the 'classical molecular gene concept,' according to which a gene is a stretch of DNA encoding a functional product, which may be a single polypeptide or RNA molecule, has been recently challenged by a series of findings (e.g., split genes, alternative splicing, overlapping and nested genes, mRNA edition, etc). The debates about the gene concept have important implications on biology teaching, and, thus, it is important to investigate whether and how these are addressed in this context. In this paper, we report results of an investigation relating to the treatment of genes in higher education cell and molecular biology textbooks. These results indicate that, despite several findings challenging time-honored ideas about genes, these ideas continue to be widely used in textbooks, even though the textbooks themselves discuss part of those findings. Textbooks also harbor a proliferation of meanings about genes that may make the concept look vague and confused, and even lead to ideas that are at odds with our current knowledge about genomes.

KEYWORDS: Biology teaching, gene, higher education, textbooks.

Introduction

The term "gene" was created by the geneticist Wilhelm L. Johannsen in 1909 and has certainly been a landmark in the history of biology in the 20th century. Snustad, Simmons, and Jenkins (2005, p. 381) argue that "the gene is to genetics what the atom is to chemistry" (see, also, Keller 2000, 2005). Indeed, few would disagree that the gene was (and still is) the central unifying theme of genetics, even though it is, at present, a "concept in tension" (Falk, 2000) or "in trouble" (Keller, 2000). This situation results from a series of findings during the last three decades that posed important challenges to our usual understanding of the gene concept, such as, split genes, alternative splicing, overlapping and nested genes, mRNA edition, alternative translation modes, etc. (see, e.g., Fogle, 1990, 2000; Falk, 1986, 2000; Pardini & Guimarães, 1992; Griffiths & Neumann-Held, 1999; Keller, 2000; Moss 2001, 2003; El-Hani, Queiroz, & Emmeche, 2006; El-Hani, 2007; Neumann-Held & Rehmann-Sutter, 2006).

In the philosophical literature, the usual way of understanding 'genes' has been designated as the "classical molecular gene concept" (Griffiths & Neumann-Held, 1999; Stotz, Griffiths, and Knight, 2004). According to this concept, a gene is a stretch of DNA encoding a functional product, which may be a single polypeptide or RNA molecule. This concept exerted a particularly strong influence on biological thought in the last century. It treats the gene as an uninterrupted unit in the genome, with a clear beginning and a clear ending, performing a single function. Several findings have shown that this concept is insufficient to grasp the complexities of the structure and dynamics of genomes. The problems with this concept of gene can be explained as a consequence of three features established by molecular biology and genetics: (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, such as those that take place in the generation of antibody diversity in the immune system); and (iii) lack of correspondence between DNA segments and RNAs/polypeptides (e.g., mRNA editing) (El-Hani, 2007). Split genes and alternative splicing are two good examples of phenomena that challenge this gene concept.

The classical molecular concept ascribes clear borders to genes, on the grounds of a relationship between a transcription unit and the sequences necessary to make a polypeptide. Split genes and alternative splicing bring considerable trouble to the idea that a gene might be such a well-demarcated structural and functional unit in the genome. Split genes contain both coding regions – exons – and non-coding regions – introns. Introns are excised during RNA splicing, in which exons are combined to form a mature, functional mRNA. In this case, the sequences transcribed into RNA are not the same as those later translated into proteins. This process poses a first problem to the classical molecular concept, which relies on the transcription unit to demarcate what is a gene. A protein coded by a spliced mRNA molecule exists as a chromosomal entity only in potential (Keller, 2005). The situation becomes more perplexing when considering alternative RNA splicing, and, consequently, the diversity of splicing patterns of the same primary transcript.

Most genes in eukaryotes contain multiple introns, and their presence allows the expression of multiple related proteins (isoforms) from a single stretch of DNA by means of alternative splicing (*e.g.*, Black, 2003). This phenomenon makes the whole idea of genes as units (no matter if structural, functional, or informational) look very clumsy. Alternative RNA splicing requires that conceptualizations of genes move far beyond the simple scheme captured in formulas, such as, 'one gene-one protein or polypeptide.' This challenge might be in principle accommodated by simply replacing this formula by a new one, such as, 'one gene-many proteins or polypeptides.' But the situation is not so simple. First, since genes do not autonomously 'choose' splicing patterns, they would lose a substantial aspect of their specificity, with regard to the polypeptides that will be synthesized. Splicing patterns are subject to a complex regulatory dynamics, which, after all, can involve the cell as a whole (Keller, 2000). Second, given that the segment of DNA which is transcribed as one unit into RNA is translated into several distinct polypeptides, it seems impossible to avoid the question 'Which segment is, in the end, the gene?' On the one hand, the unit transcribed into a single RNA could count as a gene. That is, we might call a 'gene' that stretch of DNA which can generate dozens of different proteins. On the other hand, a gene could be that unit which is translated into one specific polypeptide. That is, we might call a 'gene' each individual spliced mRNA by assuming a 'one mature mRNA-one protein' relationship. But treating mature mRNA as the gene also brings with it a number of counter-intuitive consequences. It would mean, for instance, that genes exist in the zygote only as possibilities, and do not show the permanence and stability that are typically ascribed to the genetic material. Moreover, genes would not be found in chromosomes and, sometimes, not even in the nucleus (Keller, 2000). These are only some examples among several challenges to the classical molecular gene concept, but they are sufficient to illustrate the difficulties that this concept currently faces.

Since Watson and Crick (1953), the gene has been treated as simultaneously physical matter and information, as a chemical and a program for running life. This idea is hardly trivial. Indeed, the incorporation of an informational vocabulary in molecular biology and genetics, since the 1950s (Kay, 2000), resulted in an 'information-talk' marked by the use of terms, such as, 'genetic information,' 'genetic code,' 'genetic message,' 'signaling,' etc., which can be regarded at present as nothing more than metaphors in search of a theory (Griffiths, 2001; El-Hani et al., 2006). Despite the prospect that 'genetic information' is not a clear concept in the structure of biology, but a metaphor demanding proper formulation in the context of a yet to be developed theory of biological information, genes are frequently treated as informational units, leading to the 'informational conception of the gene' (Stotz et al., 2004).

The attempt to preserve the idea of genes as units of structure and/or function (or, also, information), despite the accumulating challenges, induced a proliferation of meanings of the term 'gene,' which has become a substantial part of the crisis of this concept. The fact that the term 'gene' has several meanings is not in itself the problem, since many scientific terms are polysemous, and this does not always entail semantic confusion. The problem is rather that the several meanings of 'gene' and their contexts of application are not clearly and sufficiently demarcated, making it very likely that ambiguity and semantic confusion follows from the polysemy of the term. Symptomatically, it is now clear for many philosophers and scientists that a careful analysis and reformulation of this central concept in biological thought is necessary (e.g., Fogle, 1990, 2000; Falk, 1986, 2000; Pardini & Guimarães, 1992; Griffiths & Neumann-Held, 1999; Keller, 2000; Moss, 2001, 2003; El-Hani et al., 2006; El-Hani, 2007; Neumann-Held & Rehmann-Sutter, 2006; Pearson, 2006). It is not that one should necessarily strive for building a single gene concept, with the hardly achievable purpose of accounting for the wide diversity of meanings of this term in the scientific literature. A gene concept need not be entirely general to serve practical purposes in genetic research, and a diversity of gene concepts may have more explanatory and heuristic power. It is only that the meaning and domain of application of each concept should be well-delimited in both scientific practice and science education (El-Hani, 2007).

In an effort to organize the variety of gene concepts, Moss (2001, 2003) proposed a distinction between gene-P (the gene as a determinant of phenotypes or phenotypic differences, with no requirements regarding specific molecular sequence, nor with respect to the biology involved in the production of the phenotype), and gene-D (the gene as a developmental resource that is in itself indeterminate with respect to phenotype). He argued that genes can be fruitfully conceived in these two different ways, but their conflation is problematic, since it leads to genetic determinism. Gene-P, on the one hand, is the "... *expression of a kind of instrumental preformationism*" (Moss, 2001, p. 87), showing its usefulness due to the epistemic value of its predictive power and its role in some explanatory games of genetics and molecular biology, such as, those involved in the analysis of pedigrees. Gene-D, in turn, plays a role in entirely different explanatory games, such as, those involved in the analysis of the correspondences between DNA sequences and polypeptide products.

In evolutionary biology, yet another influential concept is the 'evolutionary gene,' despite a series of criticisms raised against it (see Griffiths & Neumann-Held, 1999; Sterelny & Griffiths 1999). This concept was introduced by Williams (1966) and elaborated by Dawkins (1982), who defined it as being "*any stretch of DNA, beginning and ending at arbitrarily chosen points on the chromosome*" that can be treated as "... *competing with alelomorphic stretches for the region of chromosome concerned*" (Dawkins, 1982, p. 87).

It is clear from the classical molecular concept that genes have been typically treated as units, indeed, "... *the basic units (the atoms) of life*" (Keller, 2005, p. 101). Despite the wide range of problems that this treatment faces (Fogle, 1990, 2000), the picture of genes as basic units of structure and function is still deeply entrenched in the scientific community. But, the gene concept has come to play an increasingly important role in contexts other than the scientific domain, and a 'gene talk' marked by oversimplified and often deterministic views invaded the public opinion, generally without thoughtful criticism (Keller, 2000). These circumstances stress the importance of investigating how genes are addressed in textbooks, which constitute a potential source from where such views are likely to reach society at large. After all, textbook discourse about genes is a potential origin of several problems in lay understanding of genes, and their relationships to phenotypic traits. Accordingly, the debates about the gene concept have important implications on biology teaching, at both high school and college levels. What must we teach about genes in a science classroom? What are the consequences of the gene crisis for biology teaching? What is the nature of the interplay between the treatment of genes in textbooks and oversimplified views found in the media and in public opinion?

As a first step in a study relating to these issues, we investigated how ideas about genes are addressed in higher education cell and molecular biology textbooks. It is important to stress that these textbooks play a key role in the training of scientists and preservice science teachers. Despite their central importance, a search, made in the ERIC database (Education Resources Information Center. http://www.eric.ed.gov) in March 2004, resulted in only one paper reporting a study about higher education genetics textbooks (Mertens & Polk, 1980), and none about cell and molecular biology textbooks. Although interesting, this study focused on textbooks that are no longer used or are rather old editions of currently used materials. In this paper, we report an analysis of the treatment of genes and gene function in three currently widely used cell and molecular biology textbooks.

Methodology

Cell and molecular biology textbooks were selected through a survey of 80 course syllabi of 67 universities located in the 5 continents, randomly chosen in Google searches. In the first step, three of the five mostly used textbooks that were identified (Karp, 2004; Lodish et al., 2003; Alberts et al., 2002) had been analyzed. Lodish et al. (2003) and Alberts et al. (2002) were the first and second mostly used textbooks, respectively. Karp (2004) was the fifth mostly used textbook. The other textbooks were Weaver (2005) and Becker, Kleinsmith and Hardin (2002), which ranked as the third mostly used textbooks (both with n = 7), but their analysis is not part of the present study.

According to the results of the survey, Lodish et al. (2003) and Alberts et al. (2002) are used throughout the world. These textbooks were identified, respectively, in 33 and 28 syllabi of university courses in U.S., Brazil, Portugal, Spain, Argentina, Uruguay, France, Belgium, Canada (both in the English and the French-speaking regions of the country), Denmark, and Australia. Karp (2004) was less represented among the surveyed syllabi (n = 5), but the results indicated that it is used in diverse countries, such as U.S., Brazil, and Canada. In many countries, these textbooks are used in their original language (English), although it is possible to find translations of them in other national languages. For instance, in Brazil there are translations of all these three textbooks in (Brazilian) Portuguese, and, among the syllabi, Spanish translations of Alberts et al. (2002), and Lodish et al. (2003), and French translations of the three textbooks were also found.

We elaborated a protocol to generate the raw data of the study throughout the careful scrutiny of the textbooks. Ideas about genes were analyzed in all chapters that made explicit reference to them, based on techniques of content analysis (Bardin, 2000). In particular, categorical analysis (Bardin, 2000) was used. In categorical analysis, one classifies extracts of texts by using significant conceptions (i.e., categories, *sensu* Bardin, 2000). Once the several relevant passages are classified, one can count the number of occurrences of each category. To obtain and explore these categories, scientific, historical, and philosophical literature about genes was employed.

The contexts within which different ideas about genes appeared have been also analyzed. In virtue of previous results indicating the variability of understandings about genes among the several subdisciplines composing genetics and cell and molecular biology (Stotz et al., 2004), the analysis focused on contexts related to specific subdisciplines. Based on previous experience with the textbooks, the following contexts for this analysis were identified: Classical genetics; Developmental genetics; Evolutionary/population genetics; Genetics of microorganisms; Genetics of eukaryotes; Medical genetics; Molecular biology/Molecular genetics; Cell biology; Biochemistry; Cell signaling; Genetic engineering; Genomics; Introduction; History of science; Glossary.

All the analyses were performed by the same researcher (the first author of this paper), so as to contribute to their reliability. In order to increase further the reliability of the study, the other two authors critically re-examined all the analyses that have been carried out and comparisons of the results with the original textbooks were also conducted by them.

Results and Discussion

The present paper focuses on the concepts of gene presented by the textbooks and their accounts of gene function. The results are summarized in Figures 1 and 2, while the results concerning the contexts in which different gene concepts appeared in the analyzed textbooks are presented in Figures 3 to 7.

Gene Concepts

Figure 1 shows that the classical molecular gene concept was frequently used in the three textbooks, but did not prevail in any of them. In two textbooks (Alberts et al., 2002; Lodish et al., 2003), the informational conception was predominant, while, in the third (Karp, 2004), the Mendelian conception of the gene as a unit of inheritance prevailed. It is clear in Figure 1 that, generally speaking, Karp (2004) exhibits the higher frequency of occurrences of gene concepts. This can be explained by the fact that this textbook is specifically focused on exploring concepts and experiments related to cell and molecular biology.

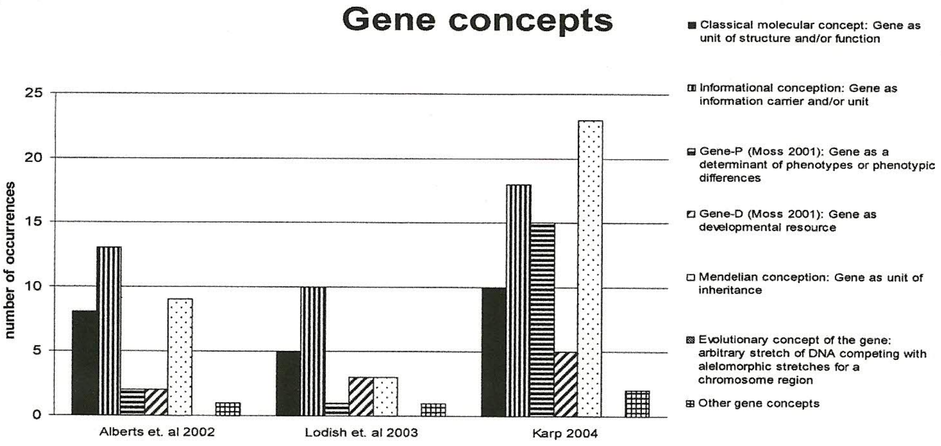

Figure 1. Occurrence of Gene Concepts in Three Widely Used Cell and Molecular Biology Textbooks.

It was somewhat surprising to identify that the molecular gene concept was not the most frequently used concept in these textbooks. Nevertheless, it was quite frequent, and, considering that the three textbooks address several findings that challenge the classical molecular concept, it becomes clear that the consequences of those findings to our understanding of genes are not properly regarded by them.

A representative example of a passage exhibiting the classical molecular gene concept is the following:

> *A **gene** is defined as the segment of DNA sequence corresponding to a single protein (or to a single catalytic or structural RNA molecule for those genes that produce RNA, but not protein)* (Alberts et al., 2002, p. 9. Emphasis in the original).

Another example, in which the authors also adopt an all-inclusive model, incorporating into the gene *cis*-acting sequences required for transcription (promoters, enhancers, terminators, regulators, etc.), is presented below:

> *In molecular terms, a gene commonly is defined as* ***the entire nucleic acid sequence that is necessary for the synthesis of a functional gene product (polypeptide or RNA)****. According to this definition, a gene includes more than the nucleotides encoding the amino acid sequence of a protein, referred to as the* ***coding region****. A gene also includes all the DNA sequences required for synthesis of a particular RNA transcript*" (Lodish et al., 2003, p. 406. Emphases in the original).

This model faces, however, several problems, mostly related to the fact that there are many different types of regulatory elements, generally operating in complex and varied combinations. There are *cis*-acting factors that influence transcription independently of their distance from the coding sequences, such as enhancers and silencers, making it difficult to empirically assign the boundaries of a gene. There are *cis*-acting factors that simultaneously affect the expression of different genes. There are even *cis*-acting factors that are nonspecific, influencing any compatible promoter within their range. Therefore, an all-inclusive model would lead to substantial overlapping of genes depending on the same regulatory sequences, raising difficulties to the idea that a gene is a structural unit. Moreover, the fact that one would be justified to include a *cis*-acting sequence in a gene, if it was merely shown that it modulates transcription leads to important problems when we consider phenomena, such as, position effects. If a rearrangement of genetic material ends up placing a gene near heterochromatin and the expression of the gene is significantly affected by this position, an all-inclusive model will demand that a whole region of heterochromatin be included in the gene. These problems, among many others, suggest that we have to abandon a completely inclusive model of the gene (Fogle, 1990). It is curious, then, that these properties of *cis*-acting sequences are at least partly discussed in cell and molecular biology textbooks, including Lodish et al. (2003), but their consequences to such a description of the gene are not taken in due account.

It is even more troublesome the predominance of the largely metaphorical informational conception, which is not accompanied in the textbooks by any generally accepted theoretically consistent account of what is biological information (after all, such an account is lacking in biology as a whole). A representative example of a passage containing the informational conception is the following:

> *In molecular terms, DNA must contain the information for the specific order of amino acids in all the proteins that are synthesized by an organism* (Karp, 2004, p. 407).

While the prevalence of the Mendelian conception cannot, in turn, be regarded as problematic, we should consider the necessity of clearly relating it to an understanding of genes that goes beyond a pre-molecular treatment of this concept. We consider, thus, that the problem with Karp is not the high frequency of that conception, particularly when we consider the contexts in which the Mendelian understanding is used (see below). Rather, the problem lies in the fact that the next more frequent treatments of genes in this textbook are aligned with the informational conception and the Gene-P concept. The former does not go beyond suggestive metaphors, as we discussed above, and the latter is committed to an instrumental, preformationist view that also characterized Mendelian genetics. Then, the task of explaining the material basis of inheritance is ultimately left to the classical molecular concept, which is also very frequent in Karp.

As a representative example of an occurrence of the Mendelian conception, we can refer to the following passage, in which we also observe an overlapping with the informational conception:

> *Gene - Physical and functional unit of heredity, which carries information from one generation to the next* (Lodish et al., 2003, Glossary, G-9).

The concept of gene-P, in turn, can be exemplified by the following extract, in which we can also detect the informational conception, and, thus, see how often in these textbooks different ideas about genes overlap in the same passage:

> *These instructions are stored within every living cell as its genes, the information-containing elements that determine the characteristics of a species as a whole and of the individuals within it* (Alberts et al., 2002, p. 191).

A number of gene concepts were quite rare in the textbooks. The evolutionary gene, in particular, was not found in any of the textbooks, but this is not surprising, since this concept seems more likely to be found in evolutionary biology textbooks. In turn, extracts that could be related to the Gene-D concept were not frequent. This is also understandable, since this idea is connected with a relatively recent approach in philosophy of biology, which does not have much currency among practicing geneticists and cell and molecular biologists, namely, developmental systems theory (see, *e.g.*, Sterelny & Griffiths, 1999; Oyama, 2000; Oyama et al., 2001). Gene-D was more frequent in Karp than in the other two textbooks, but this seems to be simply a consequence of the fact that more conceptual passages about genes were found in this textbook, generally speaking.

Ideas about Gene Function

In Figure 2, we report the results concerning ideas about gene function. In the three analyzed textbooks, the function of the gene was predominantly understood as that of codifying the primary structure of RNAs and proteins, an idea associated with the classical molecular concept. Accordingly, there is a clear relationship between the high frequency of this latter concept in the analyzed textbooks and the ideas about gene function found in them. In Alberts et al. (2002), we can find a representative example of this way of understanding gene function, when the textbook claims that "... *the genetic information in DNA is used to make both RNA and proteins*" (Alberts et al., 2002, p. 8).

Another frequent idea, particularly in Alberts et al. (2002) and Lodish et al. (2003), is the informational conception of the gene as a program or set of instructions for cell function and development. The related idea that genes control cell metabolism was also somewhat frequent in these textbooks. As a representative example, we can present the following passage:

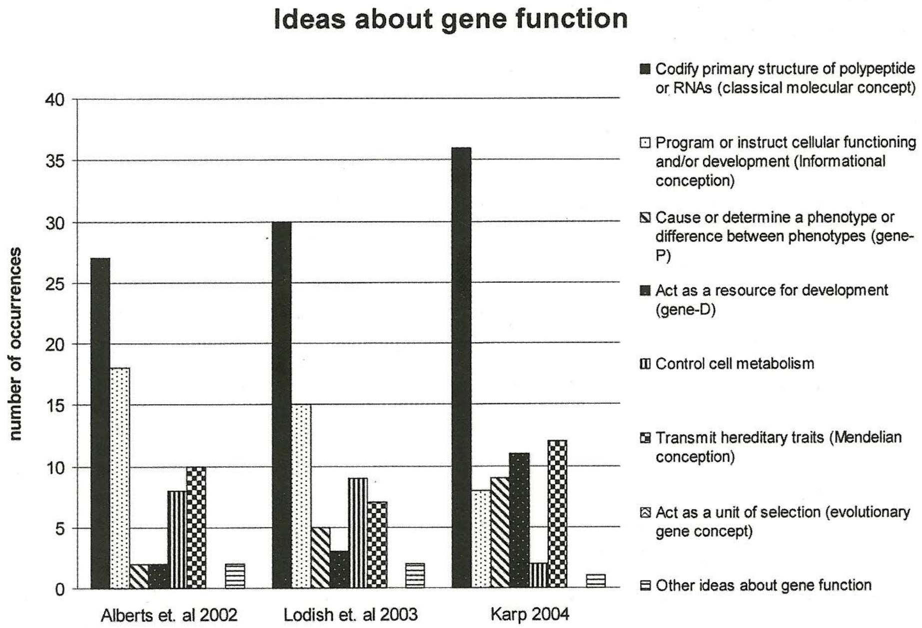

Figure 2. Occurrence of Ideas about Gene Function in Three Widely Used Cell and Molecular Biology Textbooks.

> *Although DNA stores the information for protein synthesis and mRNA conveys the instructions encoded in DNA, most biological activities are carried out by proteins. […]. For this reason, assembly of amino acids in their correct order, as encoded in DNA, is critical to production of functional proteins and hence the proper functioning of cells and organisms* (Lodish et al., 2003, p. 119).

Another representative example is found in the following passage:

> *Thus, by specifying the set of proteins that the cell is to manufacture, the genetic information recorded in the DNA sequence dictates the entire chemistry of the cell; and not only its chemistry, but also its form and its behavior, for these too are chiefly constructed and controlled by the cell's proteins* (Alberts et al. 2002, p. 12).

It should be mentioned, however, that these ideas are related to strongly criticized metaphors (*e.g.*, Oyama, 2000; Nijhout, 1990; El-Hani, 1997; Griffiths and Neumann-Held, 1999; Keller, 2000). As El-Hani, Queiroz, and Emmeche (2006) discuss, when we think of information as if DNA was a sort of reservoir from where all 'information' in a cell flows and to which it must be ultimately reduced, we tend to overplay the role of DNA in cell systems, turning it into a complete 'program for development' or an all-powerful 'controller' of cell metabolism. But, as we are enchanted by this picture of the role of DNA, we tend to forget that DNA seems to function as a set of data rather than as a program in cell systems (Atlan & Koppel, 1990). That is, we lose from sight that genes are sources of materials for cells and bookkeeping entities in evolution (Gould, 2002), playing roles that are obviously important, but cannot be correctly described as master agents (or master molecules) in cell processes (Nijhout, 1990). As Keller (2000, p. 146) suggests, if we talk about a cellular program at all, we should consider that, according to our current knowledge about cells, it is not limited to DNA, but is rather a shared program in which all cell components function alternatively as 'instructions' and 'data.' In short, we cannot neglect the fact that it is not DNA that does things to the cell; rather, it is the cell that does things with DNA. Or to use another analogy, DNA molecules are governed by the cell, rather than command the cell in a dictatorial way. Biological systems function by means of a 'democratic' rather than 'dictatorial' control structure, i.e., there is neither genomic nor metabolic supremacy over other cellular processes (Bruggeman et al., 2002).

Again, we find a relationship between a frequent understanding about genes in the textbooks, namely, the informational conception, and corresponding ideas about gene function. Also in the case of the Mendelian conception of a gene as a transmitter of hereditary traits, there is a good correlation with ideas about gene function. We can see this in the following extract, which shows also an overlapping between the Mendelian conception and the idea that genes control traits related to the informational conception:

> *The information stored in DNA is arranged in hereditary units, now known as* ***genes*** *that control identifiable traits of an organism* (Lodish et al. 2003, p. 101. Emphasis in the original).

An Analysis of the Contexts in Which Different Gene Concepts Appear

We present the data about the distribution of contexts in which different gene concepts appear in the analyzed textbooks in Figures 3 to 7, below. Figure 3 shows that the classical molecular gene concept was predominantly presented by all textbooks while addressing contents related to molecular biology and molecular genetics. This concept was also found in the glossaries of the three textbooks. It was also rather frequent in the treatment of biochemistry contents in Alberts et al. (2002) and Karp (2004). In the sections devoted to narratives about the history of genetics in Karp (2004), the classical molecular concept was also very frequent. Finally, Figure 3 also shows that this concept was found in Karp (2004) in a wide variety of contexts, including evolutionary genetics, genetic engineering, genomics, etc.

In Figure 4, we can see that the informational conception of the gene also prevailed in the treatments of contents concerning molecular biology and molecular genetics. In Karp (2004), however, this conception was more frequent in the context of cell biology. In this same textbook, there were also a relatively large number of occurrences of this conception in sections related to the history of science. It is interesting to notice that this conception was found in more diversified contexts in all three textbooks, when compared with the classical molecular gene concept. These results show how widespread this conception is, despite its lack of a solid theoretical background.

Figure 5, in turn, shows that gene-P was quite frequent in Karp (2004), in which it was mostly found in five different contexts: classical genetics, medical genetics, molecular biology/molecular genetics, and sections devoted to the history of sci-

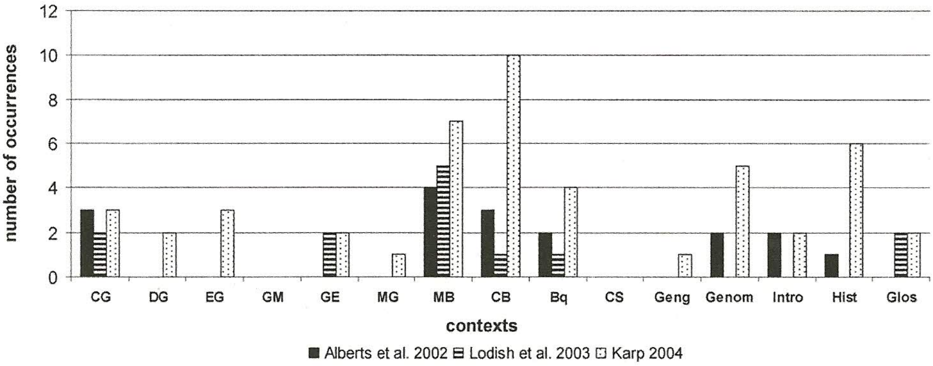

Figure 3. Distribution of Contexts in Which the Classical Molecular Gene Concept Appears in Three Widely Used Cell and Molecular Biology Textbooks. CG – Classical Genetics; DG – Developmental Genetics; EG – Evolutionary/Population Genetics; GM – Genetics of Microorganisms; GE – Genetics of Eukaryotes; MG – Medical Genetics; MB – Molecular Biology/Molecular Genetics; CB – Cell Biology; Bq – Biochemistry; CS – Cell Signaling; Geng – Genetic Engineering; Genom – Genomics; Intro – Introduction; Hist – History of Science; Glos – Glossary.

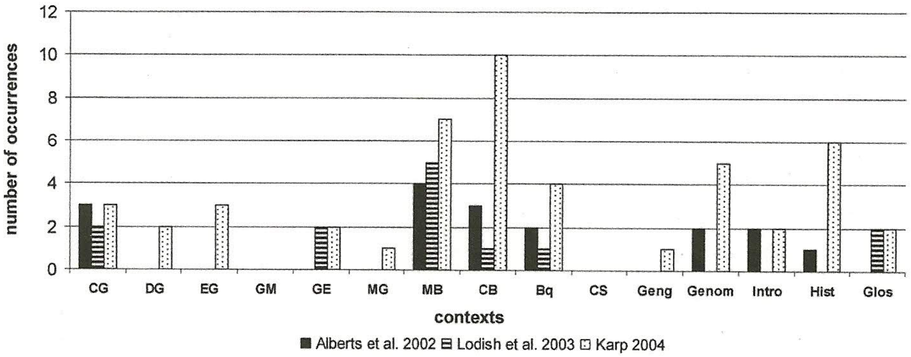

Figure 4. Distribution of Contexts in Which the Informational Conception of the Gene Appears in Three Widely Used Cell and Molecular Biology Textbooks. CG – Classical Genetics; DG – Developmental Genetics; EG – Evolutionary/Population Genetics; GM – Genetics of Microorganisms; GE – Genetics of Eukaryotes; MG – Medical Genetics; MB – Molecular Biology/Molecular Genetics; CB – Cell Biology; Bq – Biochemistry; CS – Cell Signaling; Geng – Genetic Engineering; Genom – Genomics; Intro – Introduction; Hist – History of Science; Glos – Glossary.

ence. It is consequential that the contexts in which this concept was more frequent were the sections about the history of genetics and medical genetics. In medical genetics, it is more likely that one focuses on the predictive relationship between gene loci and pathological conditions, abstracting away from the complexities of the genotype-phenotype relationship. Moreover, gene-P played quite an important role in the history of genetic research. In Alberts et al. (2002) and Lodish et al. (2003), this concept appeared in low frequency, even though its presence was

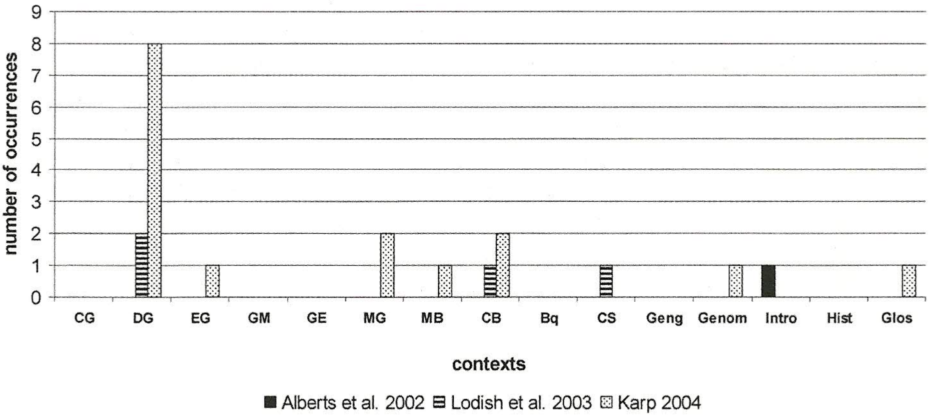

Figure 5. Distribution of Contexts in Which Gene-P Appears in Three Widely Used Cell and Molecular Biology Textbooks. CG – Classical Genetics; DG – Developmental Genetics; EG – Evolutionary/Population Genetics; GM – Genetics of Microorganisms; GE – Genetics of Eukaryotes; MG – Medical Genetics; MB – Molecular Biology/Molecular Genetics; CB – Cell Biology; Bq – Biochemistry; CS – Cell Signaling; Geng – Genetic Engineering; Genom – Genomics; Intro – Introduction; Hist – History of Science; Glos – Glossary.

detected in a wider variety of contexts in the former than in the latter textbook.

It is clear from Figure 6 that Gene-D was found mostly in Karp (2004), with a significantly higher frequency in the context of developmental genetics. This was expected, since this is a concept in which genes are treated as developmental resources. In Alberts et al. (2002), this concept was almost absent, while, in Lodish et al. (2003), it was found in three different contexts, including developmental genetics.

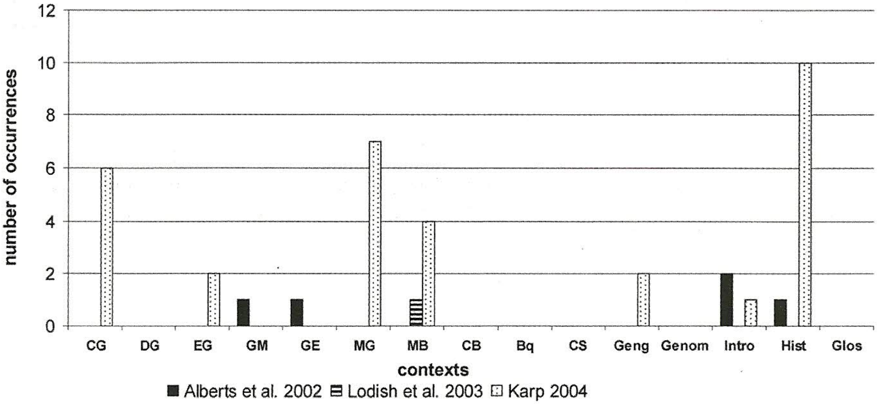

Figure 6. Distribution of Contexts in Which Gene-D Appears in Three Widely Used Cell and Molecular Biology Textbooks. CG – Classical Genetics; DG – Developmental Genetics; EG – Evolutionary/Population Genetics; GM – Genetics of Microorganisms; GE – Genetics of Eukaryotes; MG – Medical Genetics; MB – Molecular Biology/Molecular Genetics; CB – Cell Biology; Bq – Biochemistry; CS – Cell Signaling; Geng – Genetic Engineering; Genom – Genomics; Intro – Introduction; Hist – History of Science; Glos – Glossary.

We saw above that the Mendelian conception of the gene prevailed in Karp (2004). The analysis of the contexts in which this conception appears, shown in Figure 7, makes it clear that most of these appearances were in sections in which the textbook discussed the history of genetics. While this partly diminishes an initial feeling that this textbook might be committed to a pre-molecular treatment of the gene, it is still the case that the understanding of genes as material entities is left by this textbook – as we argued above – to the classical molecular concept, despite all the challenges discussed above. In the other two textbooks, the Mendelian conception appeared in a lower frequency, with most occurrences taking place, as expected, in the context of classical genetics.

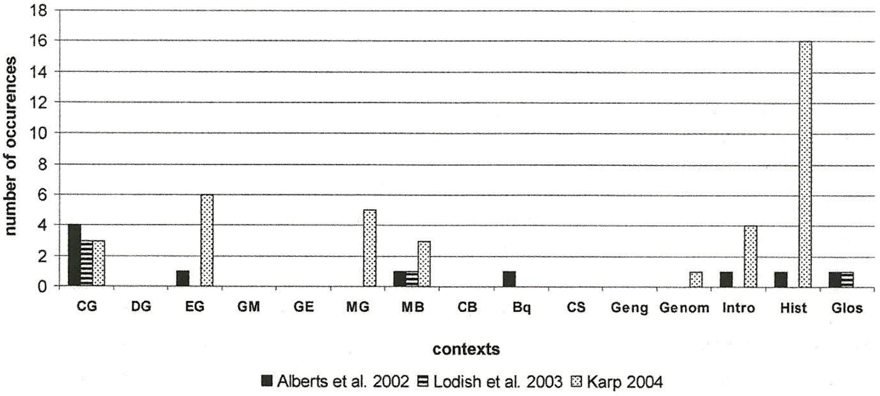

Figure 7. Distribution of Contexts in Which the Mendelian Conception of the Gene Appears in Three Widely Used Cell and Molecular Biology Textbooks. CG – Classical Genetics; DG – Developmental Genetics; EG – Evolutionary/Population Genetics; GM – Genetics of Microorganisms; GE – Genetics of Eukaryotes; MG – Medical Genetics; MB – Molecular Biology/Molecular Genetics; CB – Cell Biology; Bq – Biochemistry; CS – Cell Signaling; Geng – Genetic Engineering; Genom – Genomics; Intro – Introduction; Hist – History of Science; Glos – Glossary.

Other gene concepts occurred only four times in the three analyzed textbooks – two of them in Karp (2004) –, in the following contexts: Molecular biology/molecular genetics; genomics; evolutionary/population genetics; history of science. Generally speaking, Karp was the textbook presenting gene concepts in a wider variety of contexts. It is also clear from our results that the classical molecular concept and the informational conception of the gene are present in a larger number of contexts than other ideas about genes.

Concluding remarks

The results of the present study are interesting in several ways. First, it should be noted that although recent discoveries challenge the classical molecular gene concept and the informational conception has an unclear meaning in the current structure of biology, both are widely used in the analyzed higher education cell and molecular biology textbooks, in a wide diversity of contexts. This is curious, since these very same textbooks usually address many experimental results that led to the so-called crisis of the gene concept. Even though these results lend support to a rejection of the ideas that genes are units of structure, or function, or information (Fogle, 1990; El-Hani, 2007), these textbooks do not use them as a basis for challenging those ideas, or, generally speaking, for proposing any explicit discussion about the difficulties currently faced to understand what is a gene. Indeed, the results reported here show that there are not many attempts to address debates about the meaning of 'gene' in these textbooks.

It is also observed in the textbooks a proliferation of meanings of the term 'gene,' which may make the concept look vague and confusing, and even lead to ideas that are at odds with our current knowledge about genomes. Here, one faces the problem of assessing the consequences of conceptual variation, which has been a feature of the understanding of genes throughout the history of genetics and molecular biology. As one can see in the case of the scientific community (Stotz et al., 2004), it is observed in the textbooks a tendency to use different gene concepts in different subdisciplines or contexts. What can we say about the consequences of such conceptual variation?

Some authors remind us that this variation has been heuristically useful in the past (e.g., Kitcher, 1982; Falk, 1986; Griffiths & Neumann-Held, 1999). Consequently, the problem does not concern only the proliferation of meanings of the term 'gene.' If the context of application of each meaning was sufficiently demarcated, in such a manner that ambiguity and semantic confusion was not very likely, there would be no problem with the fact that the term 'gene' has several meanings. We have reasons to think, however, that this is not the situation, both in the scientific community and in the analyzed textbooks. Several authors warn us about the possibility that current conceptual variation with regard to the term 'gene' can lead to serious difficulties, bringing us "*... dangerously near to misconceptions and misunderstandings*" (Falk, 1986, p. 173). Fogle (1990, p. 350), in turn, argues that, "*despite proposed methodological advantages for the juxtaposition of 'gene' concepts it is also true [...] that confusion and ontological consequences follow when the classical intention for 'gene' conjoins a molecular 'gene' with fluid meaning*". El-Hani (2007) advocates that it is not necessary to propose a single, overarching gene concept, attempting to put under its umbrella all the diversity of meanings connected with this term; rather, what we may need is a diversity of definitions for 'gene' with well-delimited domains of application.

In the analyzed textbooks, the potential problems in the way genes are addressed can be generally described in terms of the absence of a systematic and theoretically-consistent treatment of the gene concept, which might make biologists in training acquire a clear and philosophically informed understanding of this central idea in biological thought. Thus, they could be able to ascribe diverse meanings to the term 'gene,' as their epistemic practices may demand, with a clear understanding of how each one of them fits into the general theoretical schemes of their science. It is not very likely that students learning about genes from these textbooks will be really able to build such a consistent understanding.

Acknowledgments

We are indebted to the Brazilian National Council for Scientific and Technological Development (CNPq) and the State of Bahia Research Support Foundation (FAPESB) for support during the development of the research reported in this paper. We also thank Aline Machado and Vinicius Casais for their involvement with the project, and Pierre Clément and Nicos Valanides for the support during the preparation of the paper.

References

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

BARDIN, L. (2000). *Análise de conteúdo.* Lisboa: Edições 70.

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

BRUGGEMAN, F. J., WESTERHOFF, H. V., & BOOGERD, F. C. (2002). Biocomplexity: A pluralist research strategy is necessary for a mechanistic explanation of the "live" state". *Philosophical Psychology, 15,* 411–440.

DAWKINS, R. (1982). *The extended phenotype.* Oxford: W. H. Freeman.

EL-HANI, C. N. (1997) Explicações causais do desenvolvimento: São os genes suficientes? *Cadernos de História e Filosofia da Ciência, 7,* 121–167.

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

EL-HANI, C. N., QUEIROZ, J., & EMMECHE, C. (2006). A semiotic analysis of the genetic information system. *Semiotica, 160,* 1–68.

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, and H.-J. Rheinberger (Eds.), *The concept of the gene in development and evolution,* (pp. 317–348). Cambridge-UK: Cambridge University Press.

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, and H.-J. Rheinberger (Eds.), *The concept of the gene in development and evolution,* (pp. 3–25). Cambridge-UK: Cambridge University Press.

GOULD, S. J. (2002). *The structure of evolutionary theory.* Cambridge-MA: Harvard University Press.

GRIFFITHS, P. E. (2001) Genetic information: A metaphor in search of a theory. *Philosophy of Science, 68,* 394–403.

GRIFFIHTS, P. E., & NEUMANN-HELD, E. (1999). The many faces of the gene. *BioScience, 49,* 656–662.

KAY, L. E. (2000). *Who wrote the book of life? A history of the genetic code.* Stanford: 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.

KITCHER, P. (1982). Genes. *British Journal for the Philosophy of Science, 33,* 337–359.

MERTENS, T. R., & POLK, N. C. (1980). A comparison of thirteen general genetics textbooks. *American Biology Teacher, 42,* 275–279.

MOSS, L. (2001). Deconstructing the gene and reconstructing molecular developmental systems. In S. Oyama, P. E. Griffiths, and R. D. Gray (Eds.), *Cycles of contingency: Developmental systems and evolution,* (pp. 85–97). Cambridge-MA: MIT Press.

MOSS, L. (2003). *What genes can't do.* Cambridge-MA: MIT Press.

NEUMMAN-HELD, E., & REHMANN-SUTTER, C. (2006). *Genes in development.* Durham: Duke University Press.

NIJHOUT, H. F. (1990). Metaphors and the role of genes in development. *Bioessays, 12,* 441–446.

OYAMA, S. (2000). *The ontogeny of information: Developmental systems and evolution* (2nd Ed.). Cambridge-UK: Cambridge University Press.

OYAMA, S., GRIFFITHS, P. E., & GRAY, R. D. (2001). *Cycles of contingency: Developmental systems and evolution.* Cambridge-MA: MIT Press.

PARDINI, M. I. M. C., & GUIMARÃES, R. C. (1992). A systemic concept of the gene. *Revista Brasileira de Genética, 15,* 713–721.

PEARSON, H. (2006). What is a gene? *Nature, 441,* 399–401.

SNUSTAD, P., SIMMONS, M. J., & JENKINS J. B. (2005). *Principles of genetics* (4th Ed.). New York: John Wiley & Sons.

STERELNY, K., & GRIFFITHS, P. E. (1999). *Sex and death: An introduction to the philosophy of biology.* Chicago: The University of Chicago Press.

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

WATSON, J. D., & CRICK, F. H. C. (1953). A structure for deoxyribose nucleic acid. *Nature, 171,* 737–738.

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

Textbooks

ALBERTS, B., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K., & WALTER, P. (2002). *Molecular biology of the cell* (4th Ed.). New York: Garland.

BECKER, W. M., KLEINSMITH, L. J., & HARDIN, J. (2002). *World of the cell* (5th Ed.). San Francisco: Benjamin-Cummings.

KARP, G. (2004). *Cell and molecular biology: Concepts and experiments* (4th Ed.). New York: John Wiley and Sons.

LODISH, H., KAISER, C. A., BERK, A., KRIEGER, M., MATSUDAIRA, P., & SCOTT, M. P. (2003). *Molecular cell biology* (5th Ed.). New York: W. H Freeman.

WEAVER, R. F. (2005). Molecular Biology (3rd Ed.). Boston: McGraw-Hill.