Against "Genes For": Could an Inclusive Concept of Genetic Material Effectively Replace Gene Concepts?

Richard M. Burian and Kostas Kampourakis

1 Introduction

In an earlier paper one of us (RMB) argued that genetics is committed to several fundamental questions, involving the localization, nature, and structure of genes, their physiology (what sorts of entities they are, what material they are made of if they are made of matter at all, what molecules they interact with, how their expression and interactions are regulated), how they influence organismal development, how they affect evolution, and how they are altered in the course of evolution (Burian 2000). According to the fundamental argument of that paper Mendelian genetics was, from the very beginning, committed to several distinct research programs that could be conveniently classed under three headings. These dealt with gene function, gene localization and composition of genes, and the functional organization of genes (meaning the functional pieces of which they were built and the relations of genes and parts of genes to one another). The case studies sketched in that paper dealt with work completed before 1940 and belong to a period described as Mendelian genetics (see also Burian 2013; Kampourakis 2013; Jamieson and Radick this volume).

We argue in this chapter that recent developments show that scientists dealing with different problems or working in different disciplines have distinct concepts of genes, but that the discrepancies in their usage (about the boundaries of genes, their precise

R.M. Burian (✉)
Department of Philosophy, Virginia Polytechnic Institute
and State University, 220 Major Williams, Blacksburg, VA 24061, USA
e-mail: rmburian@vt.edu

K. Kampourakis
Secretariat of Educational Research and Development, Geitonas School,
Vari Attikis, 16602, Greece
e-mail: konstantinos.kampourakis@gmail.com; kamp@geitonas-school.gr

localization, and the sorts of biological roles and functions that can be assigned to them) can be readily understood by understanding the disparate roles played by the genetic material (DNA and sometimes RNA) in different contexts and on different scales. We argue quite generally that in courses for students who are not planning to make professional use of genetics, rather than starting with genes it is more helpful to teach about the ways in which geneticists handle the fundamental questions about the structure of the genetic material. That is the key to understanding the differences in the claims that scientists and lay people make about what genes are and what they do. The ways in which the genetic material behaves (and the products that it yields) in different biological contexts support conflicting claims about what functions a given portion of the genetic material may have and about what its effective structure is in different contexts. Understanding this also clarifies how geneticists and other biologists test and correct their views in the light of new evidence and go about gathering and evaluating evidence which is perhaps more important than to teach specifics of highly developed models of the gene[1] or of gene action. It also helps to explain why there are continuing disagreements about what, exactly, to count as a gene and about the powers of genes. We do not believe that talk about genes is wholly dispensable, but that understanding the behavior of the genetic material is the fundamental basis for understanding the terminology involved and the continuing disputes about the nature of genes and the extent to which they "control" the traits of organisms.

By the second decade of the twentieth century, when Mendelian genetics was well established, its adherents had embarked on significant research programs. These programs fall, more-or-less, into three groups that survived and continued into the molecular era.

1. *Understanding gene function*: Insofar as Mendelian genes were defined, they were defined by use of regularities concerning the inheritance of phenotypic traits. Early Mendelian research sought to characterize genes in terms of their functions and/or consequences, together with the patterns of inheritance that they exhibited. This is the source of descriptions of 'genes for X' (e.g., eye color, height, or amount of sugar in the kernels of corn, for modifying the action of other genes, etc.). Thus, seeking to understand gene function or gene action was the core of one research program.

2. *Determining gene composition and localization*: Insofar as Mendelians were materialists regarding genes (some were right from the start, some weren't until after the Watson-Crick structure of DNA was published when most became

---

[1]Etymologically, the term 'gene' originated in the (Hippocratic) idea of Pangenesis, advanced by Darwin (1868). Darwin thought that gemmules from all parts of the body are transmitted to the reproductive organs and, from there, to the next generation. Hugo de Vries suggested that pangenesis did not involve transportation of gemmules between cells; rather all specification of information was intracellular (de Vries 1910/1889). He called the hereditary elements 'pangens', occasionally written 'pangenes' in English. It is from this that the term 'Gene' was suggested by Wilhelm Johannsen.: 'it appears simplest to isolate the last syllable, gene, which alone is of interest to us […] The word gene is completely free from any hypotheses'. (Keller 2000, p. 2, quoting from Johannsen 1909).

materialists), one needed to know WHAT a gene is, meaning 'what are genes made of?'. This was also associated with the question 'where (within the cell) are the genes located?', a key question that led to the success of the chromosomal theory. (This question has not totally disappeared in molecular genetics; it is still crucial in seeking to locate regulatory genes and to delimit the boundaries of protein-encoding genes.)

3. ***Understanding gene structure***: A final group of research programs, intertwined with (but partly independent of) the second group concerned gene structure. In particular, especially in molecular genetics, it sought to understand how the features of genes correlated with the phenotypes they produced, that is how gene structure shed light on gene action or function.

Notice that the question about what genes are made of does not necessarily answer the question of gene structure. The structure question asks not only what genes are made of but also how they can store and transmit some kind of information (for information in biology see Marcos and Arp, this volume) and the ways in which they can (and cannot) determine the traits of organisms. Once the question of how genes store information (and what sort of information they store) was solved, the structure question became more prominent and then was greatly amplified by the discoveries of split genes, promoters, enhancers,[2] and other sorts of 'elements' that modify the likelihood of readout, the stopping point of readout, the speed of readout, the combinations of genetic material actually read out, etc. Again, the question about where genes are did not end with the molecular era. The questions of where regulatory genes are located and how to delimit the boundaries of genes, including protein encoding genes, are still open.[3]

One way of characterizing the switch from Mendelian to molecular genetics is that with molecular genetics one *could* (though one did not have to) switch from working 'down' from the phenotype to the gene to working 'up' or 'out' from the gene (or the genetic material) to the phenotype. The fact that working this way is sometimes called 'reverse genetics' shows something about the limitations that genetics used to face, but no longer does thanks to molecularization.[4] One limitation

---

[2]Split genes are DNA sequences which consist of two kinds of sections: (a) those called exons, which are transcribed to corresponding RNA sections that are in turn translated into protein and (b) those called introns which are transcribed to RNA sections that are then excised and not translated; promoters are specific DNA sequences to which RNA polymerase binds; enhancers are specific DNA sequences to which proteins bind which facilitate the binding of RNA polymerase to the promoter (enhancers increase the transcription of genes – other DNA sequences have the opposite effect and are called silencers).

[3]We can generally distinguish between protein encoding genes, i.e. genes which are implicated in the synthesis of a particular protein molecule which is directly related to some trait or phenotype, and regulatory genes which are implicated in the synthesis of a particular RNA or protein molecule which in turn affects the expression of other genes.

[4]Working the other way round, i.e. from phenotype to gene or genetic material is described as forward genetics. In this case, one attempts to relate an observed phenotype to a DNA sequence. In reverse genetics, DNA sequences are usually altered in order to see which phenotypes are affected and in what way.

of molecular genetics is that the phenotypes that a gene – or genetic material – can deliver are all (more-or-less) molecular. In an important sense, there is no such thing as 'the gene for red eye' in drosophila – several to (probably) a few hundred genes are involved, including those encoding information required for producing the relevant red and the brown pigments, but also all those required to process the pigments in a coordinated way in just the right sets of cells for those pigments to yield eye color. Naming conventions are not so hard for immediate products, but are quite difficult for complex phenotypes. In Mendelian days, genes were "the factor that makes a difference to X" (where X names a phenotypic trait) but in the molecular world they are chunks of DNA (or, exceptionally, RNA) that (in some normative sense) "normally" encode certain specific products (or types of products?) or cause certain kinds of changes in what is read out, or cause the reading out to proceed at a different rate, etc. And one chunk of DNA might belong to one (or more) distinct genes, not only thanks to frameshift[5] in cases of distinct readouts, but also because inside the introns for one gene there are regulatory or even protein-encoding genes, some of which affect the readout of the gene within which they are embedded or of other distinct genes.

These considerations yield two points that greatly affect the morals that should be drawn for education. First, all three sorts of programs are required to fill in an adequate account of the gene concept. Second, the findings about the molecular structure and functions of the genetic material point to different directions – directions that may, in the end, break the concept of the gene into pieces or leave us with a much altered concept of the gene. Thus, genes are *not* contiguous pieces of genetic material yielding one (and only one) product, they may be made of different kinds of material (RNA or DNA), they need not be on (standard eukaryotic or prokaryotic) chromosomes, they need not be the least unit of function (they often encode separable functional domains[6]), etc. Empirical findings have produced (sometimes nasty) surprises and have caused major reevaluations of previous claims about how to delimit genes. There is no end in sight to the process of obtaining findings that cause scientists to revise their account(s) of gene identity, to alter the ways in which they delimit genes, and to revise what they consider to be 'necessarily' true of genes. Tensions between the findings of the different programs continue to turn up and the problem of determining the relative importance of their findings for the 'proper' delimitation of genes is not likely to go away in the near future.

Recently, much new attention has been given to these issues (e.g., Beurton et al. 2000; Dietrich 2000; Griffiths and Neumann-Held 1999; Kay 2000; Keller 2000;

---

[5]Frameshift refers to the fact that different proteins can be produced by different readouts of the same DNA sequence. The reading frames are shifted by one or two nucleotides and thus yield entirely different amino acid sequences over the length of the genetic material in which their readouts overlap.

[6]Functional domains are segments of a protein (often encoded in single exons) that play a particular well-defined role in different contexts, e.g., attaching the protein to a membrane or facilitating the interaction of the protein with a specific signaling molecule. The fact that many proteins include exons encoding distinct functional domains and that those domains are separated by introns facilitates the evolutionary process by allowing the modular swapping, addition, or subtraction of pieces performing particular functions or subfunctions.

Morange 1996, 2000, 2001; Moss 2001, 2003; Neumann-Held 1999, 2001; Portin 2002; Sarkar 1998; Snyder and Gerstein 2003; Waters 2000). In this chapter we describe various gene concepts proposed since the early twentieth century and the relevant problems in accurately defining what a gene is. Our conclusions are based on contemporary findings arising from the impact of evolutionary, developmental, genetic and medical research on the delimitation of genes and on the consequences of gene expression, plus some issues concerning public communication. We conclude that the most appropriate way of describing current genetic findings severely limits and circumscribes the use of locutions that enhance intuitive notions of genetic determinism. On the basis of these considerations, we suggest that the more inclusive concept of "genetic material" should replace the notion of the "gene" in general education about the findings of molecular genetics and allied disciplines and that it can do so effectively.[7]

2   The Gene Concept of Mendelian Genetics

The main elements of the classical chromosomal theory of the gene were fairly well established with the publication of *The Mechanism of Mendelian Heredity* by T.H. Morgan and his coworkers in 1915 (Morgan et al. 1915). According to this theory, the term 'gene' refers to a segment of a chromosome which, when activated or deactivated, performs a certain function or has a characteristic effect. But how much of a chromosome? And what functions or effects? Much of the effort that went into mapping genes may be viewed as an attempt to answer the first question; much labor was expended on the determination of which part of which chromosome contained which genes. In the process, certain criteria were developed for telling one gene from another. According to one of these, if two mutations affecting the same phenotypic trait – say two eye color mutations – could be separated by recombination, then they belonged to separate genes; if they could not be so separated, then they belonged to the same gene, and they were counted as alternative forms (alleles) of the same genetic locus.[8]

This way of individuating genes was proposed by Sturtevant (Sturtevant 1913a, b), who suggested that two closely linked eye color mutations (called 'white' and 'eosin') that Morgan and Bridges had been unable to separate in an experiment using 150,000 flies (Morgan and Bridges 1913) should be considered to be two alternative abnormal alleles of a single gene at a specified locus on the X chromosome. Now the more closely two genes are linked, the more difficult it is to separate them by recombination, and the larger the number of flies that must be used to execute the test. Thus, it should be no surprise that such claims are sometimes wrong and

---

[7]Some sections of this chapter draw in part on chapters 7 and 9 of Burian (2005).

[8]In sexual reproduction, recombination (exchange of segments of some specific length at matching loci by means of a mechanism called 'crossing over') occurs between homologous chromosomes which pair during meiosis, yielding chromosomes that have partially maternal and partially paternal DNA.

that it was established many years later, in this very case, that one can separate the two genes in question if one performs a truly gigantic recombination experiment (cf. Carlson 1966, p. 64; Kitcher 1982, p. 351).[9]

Consider the problem this creates when one asks what is referred to by subsequent uses of such terms as 'the gene for white eyes' or 'the eosin locus'. If one conforms to the usage established on the basis of Sturtevant's results, one refers to that portion of the chromosome that contains both the white and the eosin genes. But if one is working with the recombination criterion for theoretical purposes, one may refer, instead, to the smaller portion of the chromosome containing one, but not both of these genes. This is to say that two rather different segments of the chromosome belong to the reference potential (a term introduced by Kitcher 1978) of these phrases. Very often it makes no difference which portion of the chromosome one refers to, as they are, after all, virtually inseparable by ordinary techniques. But occasionally it may matter whether one purpose or the other dominates one's usage – conformity to established usage in order to accomplish coreference with other scientists or correct application of the criteria separating genes from one another. For a long time, the ambiguity was inescapably built into the mode of reference which was available in discussing these genes.

Indeed, at various stages in the history of genetics, it became a theoretical and practical necessity to distinguish between different gene concepts each of which picked out different segments of the chromosome or employed different criteria of identity for genes. For example, in the 1950s Seymour Benzer pointed out that many geneticists had assumed that the smallest unit of mutation with a distinct functional effect coincided with the smallest unit of recombination – and he performed some elegant experiments that showed that this claim is false (Benzer 1955, 1956, 1957). As a result, in some circumstances it became necessary to choose between the unit of function (which, for reasons that need not concern us, Benzer called the *cistron*), the unit of mutation (which he called the *muton*), and the unit of recombination (which he called the *recon*). This particular result showed that there had been hidden openness in the reference potential of the term (and the concept) 'gene' and that, in some arguments, though not in general, it was necessary to divide the reference of that term (concept) according to the separable modes of individuating genes.

The actual history is, of course, much richer than we have let on here, particularly when one pursues the story into the present, where one encounters transposable control elements, parasitic ("selfish") DNA, split genes with separately movable subunits, and so on. Thus, there are at least four ways in which the reference of a particular use of the term 'gene', or one of its cognates, might be specified (compare the discussion in Kitcher 1982, p. 342 ff.). Which one of these is relevant will depend on the dominant intention of the scientist and the context of the discussion.

---

[9]Carlson (1966, chap. 8) discusses the conceptual importance of Sturtevant's analysis which provided the key step in recognizing that mutation often involves alteration rather than loss of genes.

One such intention is conformity to conventional usage. Taking Sturtevant's early experimental results for granted, conformist usage would refer to the same segment of the X chromosome whether one spoke of the white or the eosin locus. Another, sometimes conflicting, intention is accuracy in the application of the extant criteria for identifying the relevant kinds or individuating the individuals of those kinds. When accuracy is the dominant intention, 'white' and 'eosin' refer to different segments of the chromosome. From this perspective, Sturtevant's 'mistake' expanded the reference potential of the term 'gene' by adding a compound chromosomal segment to the items potentially referred to by that term. In some, but only a few, contexts it proved terribly important to take the resultant long-unrecognized ambiguity of reference into account in order to understand the actual use of the relevant terms and to reconcile conflicts between competing descriptions of the outcomes of experiments. What is at stake here is the precise roles that one's theoretical presuppositions and accepted experimental results play in fixing the reference of one's terms. Although this discussion has not provided a general resolution of that difficult problem, it has given some indication of the proper apparatus to employ in carrying out case by case analyses.

The Benzer case illustrates another way in which reference may be fixed: once an ambiguity (such as that between 'cistron' and 'recon') becomes troublesome, it is sometimes necessary to stipulate as clearly as possible which of the available options one is taking as a way of specifying the reference of one's terms. Even at the risk of total failure to refer – which might happen if one's analysis is mistaken – one fixes one's reference to all and only those things which fit a certain theoretical description. The result is clarity, and when clarity is the dominant intention, reference is fixed by the relevant description. The sense of a term is determined by a description, and reference depends on whether or not anything, in fact, fits that description. Finally, one may operate with a dominant intention, which Kitcher (1978) calls naturalism, to wit, the intention to refer to the relevant effective natural kind occurring or operating in a certain situation or in a certain class of cases. It seems that one must have recourse to naturalism over and above conformity, accuracy, and clarity in order to put forth a successful account of the grounds on which Mendel, Bateson, Morgan, Benzer, and all the rest may be construed as employing concepts referring to the same thing – the gene.

3   Mendelian and Molecular Genetics

A considerable amount of laborious but fascinating experimental work during the period described as classical genetics, resulted in significant revision of Mendelian genetics. The 'pure' Mendelian concept of the gene, which was 'atheoretical' in the sense that it made no specific commitments about 'what a gene is' other than that it determined specific hereditary traits inherited in a specific pattern, was replaced by a series of improved successors which can be grouped under the label *transmission*

genetics.[10] These successors were committed to the locations of genes and gradually became committed to restricted accounts of the material of which genes were composed – roughly the protein or nucleoprotein (or some portion thereof) contained at the locus on a chromosome within which the gene was located. This extended process, both in its theoretical and its empirical aspects helped prepare the way for the advent of molecular genetics. For present purposes, we may mark that advent of molecular genetics by the identification of DNA (and RNA) as the genetic material and the publication of the justly famous solution of the principal structure of DNA (Watson and Crick 1953).

It is useful to comment briefly about the relationship between Mendelian and molecular genetics. As the reference of the term 'gene' became more tightly specified during the development of transmission genetics, in a large range of central cases the concept of the gene became that of a minimal chromosomal segment (or perhaps some compound or material within that segment) performing a certain function or causing a certain effect. The relevant effect was known as the (primary) phenotype of the gene and was essential to the identification of the particular gene in question. Not surprisingly, a major part of the history of the gene, not addressed here (see Burian 2000, 2013), concerns the interplay between what one counts as genes and how one restricts or identifies the phenotypes which can be used to specify individual genes. But when all this is said and done, a great variety of phenotypes can legitimately be used to single out genes. In this context, the reference of the concept of the gene depended on the range of phenotypes investigated. Thus, geneticists interested in improving breeds of plants and animals identified genes with effects on desirable traits (such as adult weight for meat animals and flower shape for garden plants) that could not be biochemically characterized. Such genes were not acceptable to biochemical geneticists, who required a definite identification of the biochemical differences between different gene products before they admitted differences, even if they were inherited, to count as the effects of gene differences. In contrast, evolutionary geneticists came to accept changes in the nucleotide sequence as changes in genes even when they had no other phenotypic effect. These mutations, called 'neutral mutations' came to play a major role in the development of evolutionary genetic theory (see Dietrich this volume); among other things, nucleotide sequences changes that do not alter other phenotypes (and thus do not affect fitness) help to provide a 'molecular clock'.[11]

Let us expand this last point. Thanks to the advances made in molecular genetics, it is now possible to examine changes in the DNA (mutations) fairly directly. In some cases, at least, it is also possible to track the effects of those changes rather exactly. It is now well known that some changes in the DNA are silent. That is, they have no effect on any other aspect of the structure, the development, or the composition

---

[10]It is not possible to review the major advancements of that period here but the interested reader may refer to Carlson (1966).

[11]Not all parts of the genome turn over at a uniform rate, either within an organism or between organisms. For this reason, the calibration of molecular clocks is tricky and imperfect, but with care they have proved to be very powerful analytical tools.

of the organism. Effectively, such changes in the genetic material do not amount to changes in the function of any gene, though, when suitably located, within a locus identified with a gene, they do constitute changes in the structure or composition of the relevant gene. Other changes in the DNA do, of course, result in changes in other features of the organism, but some of them do so in ways which, arguably, are of no importance to its structure, development, or function. For example, some so-called point mutations result in the substitution of one amino acid for another in some particular protein manufactured in accordance with the information contained in the gene in question. Many such substitutions have very drastic effects. But some of them, so far as can be told, do not significantly alter the way the protein folds and do not alter its biological activity or function. In such cases, there are strong reasons for tolerating in perpetuity important ambiguities regarding the referents of the concept of the gene or regarding which concept of the gene is deployed in context.

The reason for this is that phenotypes at different levels are of concern for different purposes. Consider, for example, medical genetics. If one is concerned with phenylketonuria (PKU) and allied metabolic disorders, the phenotypes one deals with will range from gross morphological and behavioral traits down to what turns out to be the heart of the matter – enzyme structure and function (Burian 1981–1982, pp. 55–59; Paul 1995). For medical purposes, both silent changes in the DNA and those changes with no effect on enzyme function often are not counted as mutations, i.e., as relevant changes in the relevant gene. Even though these changes occur within that segment of DNA which constitutes the gene of interest, because they have no relevant functional effects, the gene counts as unchanged. The reason for this is clear: the concept of the gene is coordinate with the concept of the phenotype. And the phenotype of concern is not defined biochemically at the level of DNA, but (if it is defined biochemically at all) at the level of protein or via some functional attributes consequent on the biochemistry of the relevant proteins.[12]

It is important to recognize that there are legitimately different interests that lead us to deal with different sorts of phenotypes. Evolutionists, for example, may be interested in the rate of amino acid substitutions in proteins or of nucleotide substitutions in DNA. That is, the phenotypes they are concerned with might be defined by amino acid or even nucleotide sequences, not protein function. Accordingly, their definitions of the phenotype and of the gene may be discordant with those of the medical geneticist. And it is not a matter of right or wrong, but simply a matter of legitimately different interests and explanatory aims. There are large and important specialized sub-communities in biological research with legitimately different interests, which lead them to deal with legitimately different phenotypes. As the examples introduced in the last few paragraphs show, there are serious cases in which there is no question but that those differing phenotypes correspond to different concepts of the gene and different criteria for individuating genes. Especially

---

[12]There are other changes within the boundaries of genes that may count as mutations in medical genetics, namely changes in regulatory segments of the gene that alter whether, when, where, or with what intensity the gene is expressed, or, in some cases, which exons encoded by the gene are transcribed and translated.

important is that a certain stretch of genetic material may belong to distinct genes, depending on which gene concepts are employed and on the ways in which that genetic material is utilized in the cells in which it is found. Significant examples are provided by overlapping genes (e.g., those rare cases, found mainly in prokaryotes, in which different proteins are produced by reading out sequences that are 'frame-shifted'). Again, this time mainly in eukaryotes, there are other cases in which the genetic material is read in opposite directions, with some area of overlap (Tycowski et al. 1996).[13]

Work in molecular genetics may well show that some contemporary attempts at establishing gene concepts are ill-founded. Indeed, we believe that there are clear cases of this sort, for example in sociobiology (cf. Burian 1981–1982), but also, much more generally, cases in which certain gene concepts will simply have to be abandoned in light of some of the findings of molecular genetics. But molecular genetics is compatible with several well-founded gene concepts in spite of their discordance. There is a fact of the matter about the nucleotide sequence and the structure of DNA, but there is no single fact of the matter about what the gene is, or about which gene some genetic material that has multiple uses belongs to. Even though their concepts are discordant, the community of evolutionists concerned with the evolution of protein sequence and the community of medical geneticists working on metabolic disorders are both employing perfectly legitimate concepts of the gene. This provides strong, concrete support for the claim that the concept of the gene is open rather than closed with respect to both its reference potential and its reference.

A dangling thread provides a moral for biologists to consider. Stadler (1954) distinguished between the "operational" concept of the gene and the various "hypothetical" concepts of the gene. Stadler is right that proper use of an operational concept can ensure conformity and protect against the pernicious effects of certain theoretical errors. But, as the example of white and eosin genes shows, operational criteria (here, specifically for the individuation of genes) are themselves theory-laden and quite often erroneous. Furthermore, there is no single operational concept (or set of operational criteria) for the gene. In the end, as the brief discussion of molecular genetics in the last few paragraphs suggests, the best arbiter we have of the legitimacy of both operational and hypothetical concepts of the gene comes from molecular analysis. The latter, in turn, cannot be extricated from what Stadler would have considered a hypothetical concept, namely that of the structure of the DNA molecule. It follows that genetic concepts (and theoretical concepts generally) are inescapably open in the ways we have been describing.

---

[13]For many other puzzling cases in which geneticists employ concepts that yield discordant answers regarding whether a gene is present or how genes should be delimited or enumerated, see Stotz et al. (2004) and Griffiths and Stotz (2006).

4   Gene Concepts

Broadly speaking, there are two kinds of gene concepts. In this section we will offer a modest account of each. Both kinds are legitimate and understanding their interplay is crucial for understanding the history of genetics and a number of current issues in genetics. The first kind of concept makes sense of the conceptual continuities in the history of genetics, but yields concepts that are too generic or schematic to specify adequately what is referred to by '*the*' gene concept and allied concepts. Without such generic or schematic concept(s) of the gene, there would be no such discipline as genetics. However, without supplementation by more specific gene concepts, the schematic concepts do not suffice for specifying the referent of the term 'gene' – indeed, they do not specify well enough what genes are to ensure that the term refers successfully at all. In less philosophical language, these schematic concepts are impotent to specify exactly what we are talking about when we talk about genes.

   The second kind of gene concept, in contrast, yields specific gene concepts, but does so at the price of conceptual discontinuity. If one restricts oneself to the series of discontinuous gene concepts, the findings of molecular genetics favor abandoning a univocal and specific concept of the gene altogether in favor of a pair of concepts – the concept of genetic material plus that of the expression of genetic information. This conceptual change allows molecular genetics to bypass the problem of discontinuity, currently solved by the use of schematic gene concepts. It also solves several other problems. As some other scholars have argued, the information content of the genetic material is extremely dependent on the cellular or subcellular context in which it is expressed.[14] This provides one of the rationales for suggesting that molecular biologists could abandon specific concepts of the gene, deploying, instead, concepts focusing on the continuous genetic material and the controls governing what is still called gene expression.

4.1   Schematic (i.e., Referentially Indefinite) Gene Concepts

Any science that seeks to locate hidden causes of some spatio-temporally delimited class of phenomena must use indefinite descriptions. These are descriptions that leave the exact referent of a term open. An example would be a Mendelian description like '*the factor, whatever it is, in the germ cells of these peas that causes them to produce plants that are much shorter than the tall plants produced from peas from the same pod*'. Such specifications are indefinite in not specifying what the causal factor in question is or even what category or sort of thing or process the factor is. Indefinite descriptions can genuinely refer to entities, as does the example

---

[14]The papers by Falk, Fogle, Gifford, Gilbert, Holmes, Rheinberger, and Schwartz in Beurton et al. (2000) are particularly relevant to this point.

we just gave when used in the right circumstances, but they can also be associated with seriously false descriptions or commitments. This is illustrated by the commitment, common before the middle of the twentieth century, that Mendelian factors (or genes) are composed of proteins. Mendelian genetics, taken strictly (i.e., without commitment to the localization of genes on chromosomes), used gene concepts based on very open-ended indefinite descriptions of exactly the form illustrated above.[15]

We call concepts like that of a gene thus understood *referentially indefinite causal (or functional) concepts*. In particular, the identification of a gene illustrated above is indefinite, but is accomplished in terms of a two part functional description. The first part specifies a difference in the phenotype of the organism bearing a gene (tall vs. short); the second requires a pattern of transmission of the factor(s) responsible for the change. One can distinguish different genes affecting, say, a plant's height or its flower color by their behavior in breeding experiments, by whether or not they 'Mendelize' or follow some recognizable variant of classic Mendelian patterns of inheritance (e.g., 3:1 or 9:3:3:1). Transmission genetics adds a third constraint on identifying genes, namely, their localization on a chromosome.

Here is a schematic formulation of a referentially indefinite functional gene concept: A gene for trait x is any stably inherited factor that causes an organism [or certain cells of the organism], given the rest of what it has in common with conspecifics, to have the potential for manifesting x, where x will (or can be made to) appear under the appropriate developmental plus environmental circumstances.[16] Distinct genes for x may exist and may be discriminated from each other either by specific differences in the phenotypes they cause or by demonstrating that they can be inherited independently of each other. This scheme instantiates Stadler's (1954)

---

[15]Johannsen's attempt at an atheoretical definition (Johannsen 1909) illustrates the point precisely. In Carlson's translation (Carlson 1966, pp. 20–22): "The word 'gene' is completely free from any hypotheses; it expresses only the evident fact that, in any case, many characteristics of the organism are specified in the gametes by means of special conditions, foundations, and determiners which are present in unique, separate, and thereby independent ways – in short, precisely what we wish to call genes." Genes are thus the differences, whatever they may be, between gametes that cause organisms to have the potential for revealing patent, independently-heritable, traits. Darden (1991) amplifies this point usefully in firmly separating Mendelian genetics as developed after the 'rediscovery' of 1900 from the chromosomal theory developed by the Morgan group and others.

[16]It is important to note that as we develop an account of the relevant causal chains, we may come to adjust what we count as a trait or, at least, what we count as a trait caused in a particular, stably inherited, manner. Think, for instance, of the multiplication of distinct disease entities, e.g., some of the cancers formerly believed to be a single disease, as we have learned to distinguish different underlying ways in which, e.g., the regulatory apparatus of certain types of cells can be disrupted so as to yield phenotypically similar outcomes. It is also important to recognize that the schematic definition may require specification in a great variety of ways. Thus the specification of 'modifier genes' (i.e., genes that have the function of altering the expression or function of other genes) and 'regulatory genes' may be relative to a specific gene or control pathway carried by some, but not all, conspecifics affecting their manifestation of the relevant traits affected by the modified gene. Again, transmission genetics requires the specification of a chromosomal location for the gene over and above the rest of its Mendelian characterization.

'operational gene concepts', indefinitely described. Two points are involved: first, for a long time there were competing theories about the material constitution of particular operationally defined genes, between which no decision was possible. Second, breeding procedures allowed workers to distinguish between distinct genes with otherwise identical phenotypic effects.

Such concepts need not imply any direct claims about what genes *are*, e.g., what they are made of; in general they do not specify the material or structure of the gene(s) in question and even in the best cases they do not, by themselves, pin down their full structure. Without independent knowledge of gene structure or composition, then, these concepts do not provide a fully adequate way of individuating genes. (That is why, in the absence of knowledge about the material composition of genes, Stadler was so pessimistic about our ability to resolve questions about 'the hypothetical gene'.) If adequate information about structure or composition of genes is not built into the gene concept or if it is not determined on independent grounds, it is not possible to count genes in a stably satisfactory way. This helps make sense of the fact that the chromosome theory – or something like it – was flatly needed to complement or complete Mendelian genetics. And it helps explain part of what is accomplished by the specification of genes as composed of DNA and RNA. Once such additional information is built into the concept of the gene, the theoretical presuppositions of gene concepts are radically strengthened – and, for most of the history of genetics, the presuppositions involved have been substantially false.[17]

One can view the history of genetics as involving, among other things, a series of attempts to obtain experimentally and conceptually sound ways of filling in indefinite descriptions of genes of the sort suggested above. What *should* count as a gene, given the indefinite starting point, depends on the specific traits or functions examined and the patterns of inheritance that they exhibit. It also depends on larger commitments, such as the means we employ to determine that something (e.g. a particular sequence of nucleotides), in context, is causally responsible for the trait differences in question. It depends, further, on the restrictions we place *in context* on the ascription of causal responsibility. In the century or so with which we are concerned, it has been at various times stoutly affirmed and stoutly denied that in order to count as a gene an entity had to be on, or to be a part of, a chromosome, or composed of protein, or composed of nucleic acid, and so on. In general, there is no adequate way of telling when such claims were intended as conceptual and when they were intended as factual claims. For this and other reasons, to make sense of the history of genetics we need to understand that *when such commitments had conceptual force, there was always a pathway of retreat open*. The underlying concepts to which people retreated when necessary were referentially indefinite functional concepts.

---

[17]This claim is, of course, contentious, but we believe it is correct. Consider the sorts of substantially false conceptual commitments that have commonly been made: genes are discrete particles, genes are composed of proteins, they are located only on chromosomes, they are linearly contiguous, they are non-overlapping, etc. Note that our claim that such commitments of detail have been built into gene concepts and are substantially false, does not imply that genetics is based on fundamental mistakes.

It should be clear that indefinite descriptions of genes, even when conjoined with massive sets of experimental results, are not sufficient to specify exactly what terms like 'gene' or 'gene for x' and their cognates refer to. One thing that is often meant by a (or 'the') theory of the gene is the theory-based specification of what it is that goes into individuating genes beyond the indefinite descriptions plus sheer experimental findings. A great deal is involved here. Among the things that should be included are abstract principles for the delimitation of causes, the delimitation of the biological functions to be examined (cf. visible phenotypes vs. behaviors vs. protein structure), and commitments about the material composition, structure, or localization of genes that constrain the concept of a gene and the possible referents of that concept. To understand the historical continuities that make genetics into a discipline and give geneticists a series of problematics on which to work, it is necessary to recognize this role of referentially indefinite concepts, but also to recognize that referentially definite concepts (or, at least, referentially more definite concepts) are requisite for specifying *what genes are* and what is needed to develop means of testing the principal claims made about them – claims about how to individuate them, how they act, and so on. The need to answer such questions has had considerable impact on the character of theory in genetics. Indeed, the failure to develop globally satisfactory definite descriptions of genes is part of what moves us to suggest the need for conceptual reform in molecular biology.

4.2 Definite Gene Concepts

More specific concepts of the gene, though they may still allow further specification, are committal, at least to some degree, about the structure or the localization of genes. What is typically required is a mixed mode of identification in terms of both structure and function. When such definite concepts embody false presuppositions they may, if taken literally, turn out not to refer to anything (e.g., when they make the mistaken commitment that genes are composed of proteins) or they may apply to a subclass of the entities currently considered to be genes in molecular biology (as do those gene concepts that require genes to be composed of DNA, which miss the genes of RNA viruses and several other relatively obscure entities that utilize RNA as their genetic material).

It is always possible to retreat to a less definite description of genes and to constrain successful use of the terms in question so that they must refer to a causal factor contributing to the occurrence of a well specified phenomenon. Of course, in principle, they might then end up referring to an integron (see Rheinberger 2000), and not DNA or RNA as such at all. Thus, it is (nearly) always possible to retreat from false presuppositions so that it is clear that the claims of scientists who employed those presuppositions made good sense (see Burian 2005 chap. 7; see also Burian et al. 1996; Kitcher 1978, 1982). But it is also true that one must specify the substrates out of which genes are built and the structures that can count as

relevant causes (and thus deserve to be identified as genes) in order to individuate genes among the thicket of factors contributing to the relevant functional state. Note that for this class of gene concepts the choice of a phenotype is crucial in determining what counts as a gene; when the phenotype is an amino acid sequence, genes will be individuated differently than when the phenotype is something like the suppression of the expression of certain other genes. And it will continue to be the case that biologists with different interests will seek genes for phenotypes of different sorts. Thus, one cannot escape the recognition that there are sharp discontinuities in the history of genetics – discontinuities that cannot be bridged directly ('genes must be composed of protein' vs. 'genes must be composed of nucleic acids'). Nonetheless, such differences can be bridged via a retreat to less definite descriptions.

Once this point is granted, it is clear that the findings of molecular biology, some of which we allude to briefly in the next section, are readily interpreted as calling into question whether genes are particulate without preventing those of us who deny that they are from referring to the same things that our forefathers in Morgan's and Bateson's groups did when they used terminology committed to particulate genes and dynamic equilibria respectively. Indeed, given our treatment of concepts, the findings of molecular biology allow us to deny that the terminology of genes is well-defined and that it picks out a well-delimited group of entities. Given the range of functions for which we seek genes, one may even doubt whether all the gene-like causes are restricted to nucleic acids (cf. prions). But let us set that issue aside so that we may deal with the question whether we have a good way of settling which parts of which DNA and RNA molecules ought to be considered to be genes in light of contemporary knowledge. To this question there seems to be no systematically satisfactory answer. The best answer in a given case depends on our purposes and on the schemes of classification we employ, both of the functions that may be caused genetically and of nucleic acid molecules and their parts.

5   Continuities in the Genetic Material or Why It Is Impossible to Structurally Individuate Genes

Within rather broad limits, we are free to use terminology as we choose. We should, of course, be clear about our usage in order to avoid the confusion that results from using preempted terms in ways that conflict with common usage. The term 'gene' in molecular biology is a genuine accordion term – its expansion and contraction have caused a great deal of semantic quibbling. But the arguments involved are sometimes substantive, for they turn on the inclusion or exclusion of a number of genetic functions performed by nucleic acids that do not fit any of the standard structural constraints on genes. Underlying the different terminologies are serious disagreements about the status of parts of nucleic acid molecules that behave or are treated in different ways in different cellular contexts and at different phases of

ontogeny. Here, for example, is one of the broadest gene definitions (specifically, of eukaryotic genes) in the literature:

> We define a [eukaryotic] gene as a combination of DNA segments that together comprise an expressible unit, a unit that results in the formation of a specific functional gene product that may be either an RNA molecule or a polypeptide. The DNA segments that define the gene include the following:
>
> 1. The transcription unit refers to the contiguous stretch of DNA that encodes the sequence in the primary transcript; this includes (a) the coding sequence of either the mature RNA or protein product, (b) the introns, and (c) the 5′ leader and 3′ trailer sequences that appear in mature mRNAs as well as the spacer sequences that are removed during the processing of primary transcripts of RNA coding genes.
>
> 2. The minimal sequences needed to initiate correct transcription (the promoter) and to create the proper 3′ terminus of the mature RNA.
>
> 3. The sequence elements that regulate the rate of transcription initiation: this includes sequences responsible for the inducibility and repression of transcription and the cell, tissue, and temporal specificity of transcription. These regions are so varied in their structure, position, and function as to defy a simple inclusive name. Among them are enhancers and silencers, sequences that influence transcription initiation from a distance irrespective of their orientation relative to the transcription start site (Singer and Berg 1991, pp. 461–462, see also pp. 435 ff. and 457 ff.).[18]

This definition includes a great deal that others would exclude. A more orthodox definition, like that of Goodenough and Levine (1974, p. 291), would restrict the gene to those nucleotides which, "when transcribed, will produce a biologically active nucleic acid," thus excluding promoter sites, enhancers, silencers, introns, and the like. But no matter: on either definition most eukaryotic genes are discontinuous stretches of continuous DNA, since introns are excised from biologically active protein-encoding RNAs. Worse yet, in many eukaryotes and quite a few prokaryotes, chain termination is dependent on physiological circumstances and/or is developmentally regulated. This means that the size of a gene – or what parts of the DNA of a multigene family function as genes rather than counting as pseudogenes – depends on physiological circumstances or developmental stage. Even worse are the cases in which RNA is edited (i.e., systematically altered by specific biological processes after it has been transcribed from a DNA source) or DNA encoding immune system proteins is systematically 'shuffled' during development in different cell lines, thus making a greater variety of immune proteins than were originally encoded in the zygote. Such shuffling of the genetic material means that the original genetic contents of a zygote (i.e., a fertilized egg) are not preserved in certain somatic cell lineages. The dynamism of the genome is of great importance for the definitional and conceptual issues that belong at the heart of this chapter.[19]

It might be thought that this argument can easily be dismissed as a trivial semantic argument about how we should define terms, rather than as an argument bearing on

---

[18]Chapters 11 and 12 in Burian (2005) contain illustrations that will help the reader unfamiliar with the technical terminology to understand Singer and Berg's text.

[19]For reviews providing some details and amplifying on the importance of such issues, see Fogle (2000) and Portin (2002).

how we should think about genes in light of the findings of molecular biology. The next argument, however, which focuses on protein-encoding genes, shows that the issues just raised are not merely semantic in this pejorative sense, but have significant impact on our interpretation of the history of genetics and impinge on how biologists and lay people should be thinking at this point.

The argument concerns the continuity of the genetic material: It yields an important intermediate conclusion: An examination of intrinsic features of RNA or DNA is not sufficient to delimit precisely which parts of these molecules should count as protein-encoding genes because of the context dependence of the "readout" produced from a sequence of nucleotides.[20] It takes an enormous amount of biological machinery for genes to be expressed; exactly which parts of the genome are processed depends on specific settings and structure of that machinery. Again, a huge number of processing steps affect the times and places at which informational molecules yield products as well as exactly what products they yield. It was known as early as 1987 that the translational apparatus alone requires some 200 macromolecules (Freifelder 1987, p. 367)! Corresponding to the richness and variability of the mechanisms involved, is the richness of the alternative results (even at the molecular level) when a given stretch of nucleic acid is transcribed or enters into an interaction of some sort. The answer to the question *which stretches of nucleic acid should count as genes* depends not only on the functions and the sequence of nucleotides that we have chosen to examine, but also on the particular machinery present in particular cells or compartments within cells, for that is what determines which parts of the signal remain intact and are contiguously read out and what the molecular results of the network of interactions involved turns out to be.

As is generally known, there is cellular machinery that determines which stretches of DNA are accessible to RNA polymerases and where it is that the RNA polymerases get stopped or knocked off the DNA (both dependent, for a given stretch of DNA, on physiological conditions), and how the resulting RNA is processed – immediately in prokaryotes and before it can get through the nuclear membrane in eukaryotes. It is worth recalling at this point, that in eukaryotes, most genes are processed in such a way that the material corresponding to introns is snipped out of the RNA molecule before the transcript gets through the nuclear membrane. At least occasionally, some of the material thus snipped out is, in turn, translated to yield a functional polypeptide or is functional in some other way (Tycowski et al. 1996; Coelho et al. 2002), so that it is natural to talk of one gene embedded inside another.[21] There may still be further

---

[20]This is also the reason for which the amino acid sequence cannot be determined (or determined up to permutations) by an examination of the structure of DNA or mRNA molecules alone. In different cellular contexts (nucleus vs. mitochondria, some species of organisms vs. others), there are sometimes some regular differences in the transfer RNAs. Thus, in a few cases, the same codon in different contexts codes for a different amino acid or for a stop signal instead of an amino acid.

[21]A brief technical description of such a case is given by Singer and Berg (1991, pp. 705–706) for introns in the mitochondria of yeast.

post-transcriptional processing of mRNA,[22] and, at that, what precise polypeptide sequence the RNA yields is still a function of the tRNAs in the relevant cytoplasmic location. Further, post-translational processing of proteins is, at least in some cases, critical to whether or not the product that results in fact enters into a final product that plays a functional role.

Perhaps a schematic example will make the point clearer. Consider an ORF,[23] located by appropriate molecular techniques. Does the ORF mark the beginning of, or even delimit, a gene? The answer, insofar as there is one, depends on the physiological context, the alternative splicing and readout controls present in the relevant cell compartment (for the stop signals are different in mitochondria than in the nucleus), the tRNAs present in the immediate context and so on and on. Often enough, a single ORF begins a transcript that contains multiple genes.[24] Our conclusion is that even when one works at the molecular level, what counts as a gene is thoroughly context dependent.

An important effort to take context into account is Lenny Moss's *What Genes Can't Do* (Moss 2003).[25] Moss distinguishes sharply between two sorts of gene concepts, labeled *gene-P* and *gene-D*. The label *gene-P* is meant to capture the connection between preformationism and genes that determine a phenotype; thus, a gene-P is defined as a gene *for* a phenotype (i.e., one that is identified by its causal link to that phenotype) (Moss 2003, p. 45). In contrast, a *gene-D* (the 'D' indicates that the gene is interpreted as a developmental resource) is defined by its molecular sequence (i.e., intrinsically, without reference to what it produces). Moss rightly insists that a nucleotide sequence may enter into many different interactions and may be processed so that the products it yields have many different structures that occur in many different tissues. Similar things may be said for non-coding (regulatory) nucleotide sequences and the reactions that they affect. Accordingly, it is simply incorrect to identify molecular sequences in terms of particular effects. No gene-D is properly understood as a gene for X, where X stands for a single phenotype or a function; the effects of a gene-D depend on the biological context and (often) on the history of the organism. Hence, the effects of a gene-D are "*indeterminate* with respect to phenotype" (Moss 2003, p. 45).

---

[22]Alternative splicing, i.e. the production of different mature mRNAs from the same primary RNA transcript through differential excision of introns, is just one of many relevant post-transcriptional phenomena that are relevant here. Gilbert (2000) and Singer and Berg (1991, pp. 578) provide helpful accounts of alternative splicing and other technicalities discussed below. This phenomenon again demonstrates the impossibility of employing the intrinsic features of the DNA or RNA alone to determine which stretches of a DNA or RNA molecule produce "biologically active RNA." For further explanation of many of the issues discussed below, see Burian 2005 chap. 12.

[23]The abbreviation stands for "Open reading frame," which is the name for the nucleotide sequence that signals (in many but not all contexts) a place at which to initiate the readout of DNA.

[24]Chapter 5 of (Gilbert 2000), which covers differential gene expression, includes useful reviews of differential RNA processing (pp. 130–133) and of (contextually variable) translational and post-translational controls of the end products of the expression of nucleotides sequences (pp. 134–136).

[25]Waters (1990, 1994, 2000) offers a contrasting approach to this problem.

This point about nucleotide sequences and the indirectness of their relationship to phenotypes is entirely correct. But we are skeptical of Moss's deployment of the terminology of genes-D. The problem is how one delimits one gene-D from another. Not all nucleotide sequences should count as genes. Some short nucleotide sequences are repeated millions of times within the genome. Should each arbitrary length of such a sequence count as a distinct gene? For good reasons, even when one is working at the molecular level, it is often desirable to identify distinct nucleotide sequences as instances of the same gene – e.g., in numerous contexts in which the relation between a gene and amino acid sequences is at stake, synonymous substitutions are counted as alterations that do not change the identity of the gene, even at the molecular level. Moss would probably consider this a confusion of gene-P interpretations of the gene with gene-D interpretations of the gene. We consider it evidence that *even at the molecular level, functional criteria of delimitation are built into gene concepts*. The issues here obviously ramify far beyond this immediate, partly linguistic, partly conceptual point. Moss's insistence that we take seriously the idea of a sequence-defined or sequence-delimited concept of the gene is salutary. The issue is over the need to restrict sequence-based definitions with further (functional) criteria in order to save the gene concept from picking out any and all arbitrary sequences. In either case, the result is that *the context-dependence of the effects of nucleotide sequences entails that what a sequence-defined gene does cannot be understood except by placing it in the context of the higher order organization of the particular organisms or subcellular units in which it is located and in the particular environments in which those organisms live*.[26] This argument provides a synopsis of the one strand of support for the claim that the science of genetics has argued itself out of the most stringent versions of reductionism.

6   From the Reductionism of Genes to the Complexities of the Genetic Material

We have not yet given a working definition of the genetic material. It is now incumbent on us to do so. Genetic material is any material that provides the information utilized in constructing (other) materials within the same cell or organism with specific biological functions. In contemporary genetics and molecular biology, the use of 'information' in this context is very special and widely misunderstood. Information in this special sense is always *sequence information* for constructing sequences in new (potentially information-bearing) molecules; so far as is now known, the constructed sequences are either sequences

---

[26] It should be noted that (to the best of current knowledge) sequences of nucleotides in plasmids, viruses, mitochondria, and plastids do not replicate or reproduce outside of the laboratory except in cellular contexts. Thus, for practical purposes, the only contexts in which these entities have functional genes are when they are in some sort of cellular context.

of nucleotides in a nucleic acid or sequences of amino acids in a polypeptide in accordance with the (contextually specific) genetic code. In principle, other materials might have similar information-bearing functions, but the only known materials with such functions are the nucleic acids DNA and RNA. This special sense of information was first proposed by Francis Crick (1958).[27]

The key point is that sequence information goes from nucleic acids to proteins. Proteins do alter nucleic acids, e.g., by, by annealing nucleic acids, cutting pieces out of them, or providing machinery for occasional substitutions of one nucleotide for another, but proteins do *not*, as such, contain or provide sequence information for determining sequences of nucleotides or sequences of other amino acids. Thus, proteins can cause alterations of nucleotide sequences, but they do not contain information for constructing specific sequences. If one understands 'information' as sequence information, it becomes clear (and remains correct) that genetics has captured an extraordinary feature of nucleic acids that is not matched by proteins. This justifies the distinction between hereditary traits that are genetic (i.e., specified by genetic information) and hereditary features that are not genetic (i.e., that are specified in other ways). But it also restricts the phenotypes that count as genetic and justifies the claim that there are also a variety of forms of non-genetic or extra-genetic inheritance, i.e., of epigenetic inheritance. Thus the pigment molecules that produce the red color of drosophila eyes are specified genetically, but that it is the *eyes* that are red is specified by developmental controls that are (in part, at least) epigenetic, for those controls determine when and where the two relevant pigments are distributed and in what proportions. The current technical definition of epigenetic inheritance is (regular, lawlike or mechanistically explained) inheritance of specific states or changes of state that do not depend only on nucleotide sequences or changes of nucleotide sequence. Cellular inheritance and organismal inheritance of methylation of nucleotides or histones, or of chromosome conformation (e.g., via histone modifications) are the easiest examples of epigenetic inheritance, but more contentious examples include behaviors of mothers (for example, grooming of rat pups that causes heritable methylation that, in turn, causes many inherited effects, including increased likelihood of grooming behavior) (on this topic, see Uller this volume; Jablonka and Lamb 2005; Jablonka and Raz 2009).

---

[27] This is Crick's version of the 'Central Dogma of Genetics'. Watson's, version was different: he interpreted the Central Dogma as claiming that information flows from DNA to RNA to protein, and not backward. Watson's formulation was especially influential thanks to the importance of his textbook *The Molecular Biology of the Gene* (Watson 1965), but it was mistaken and caused a lot of the resistance to reverse transcription (see Strasser 2006; Morange 2008; Olby 1972, 1975). Unfortunately, although in practice geneticists often use the term information in accordance with Crick's account, they often present Watson's account of the Central Dogma when they discuss it, a situation that has caused much misunderstanding, even by geneticists, of their discourse about genetic information. The confusion has been exacerbated by confusion between claims about sequence information with claims about information, in some more general sense, e.g., as it is used in cybernetics or information theory.

Not all of the genetic material is (or should be) counted as belonging to specific genes. Accordingly, conceptually speaking, what counts in classifying some genetic material as belonging to a gene depends on the genetic material in question having an effect on what is counted as a phenotype in at least some circumstances. What one may choose as a phenotype, however, is somewhat constrained by what we learn about the genetic material. Factually speaking, the delimitation of genes at the molecular level depends on the entire system for processing DNA and RNA, the translation of processed RNA into protein or into regulatory products, and also the post-transcriptional processing of those products and the post-translational processing of proteins. As a result, the task of delimiting genes contains an inextricable mixture of conceptual and factual elements. To be sure, the 'lowest' level', i.e., the molecular level, though it is most distant from naive observation, brings the argument closer to a context-fixed factual basis than the others. But the price for this is that one must deal with the interactions of all of the relevant macromolecules and regulatory elements within their physiological setting to tease out the more narrowly delimited specific definitions of genes and gene functions. This has the consequence that precise definitions of genes must be abandoned, for there are simply too many kinds of genes, delimited in too many ways for a single characterization to work. Taken in combination, these arguments provide powerful support for the principal contention of this chapter, namely that when we reach full molecular detail we are better off to place careful limits on specific gene concepts.

Since the 1980s, with the advent of genomics, high throughput databases, and the many other technological and experimental advances fostered by the Human Genome project, serious work in molecular phylogeny and comparative and technical studies at the molecular level have brought about a revolution at a foundational level of our understanding of genes, genetics, and genomes. Molecular and bioinformatic tools have enforced reorganization of our knowledge and what we used to consider as solidly established findings about genes became contextually limited or approximate truths. This revolution is largely quiet; although a lot of the details are familiar, they have seemed fairly particular and the large-scale changes that they will almost surely bring in their wake have remained largely undigested and have not yet been assimilated into wider public consciousness. This revolution is ignored in the medical world (at least as understood by the larger public) to the extent that the Holy Grail that is (all too often) sought there is "the gene for". In fact, what is typical, and what quite a few researchers have cottoned to, is that researchers seek to identify key steps in various physiological process that are controlled by some product of some gene in rather particular contexts. Worse yet, it is also widely recognized that in most interesting cases, there are several networks of various sorts (gene networks, protein networks, physiological process networks, and networks that have nodes of all these sorts of entities) that intersect in controlling or contributing to the disease or processes of medical interest (Goh et al. 2007).

Most eukaryotic genes do not have very well defined boundaries. If one looks at the standard definitions of a protein encoding gene, what one gets back is a mixed bag that amounts to this: what counts as a gene is the largest unit that corresponds to a member of a family of proteins (such as one of the myosins),

and can be read out in various ways, differentially in different tissues, to yield different members of that family. In general, this is NOT the largest unit that can be read out from the same start site since about 0.5 % of readouts do not end at the standard stop signals but contain material from two, three, four, or more conventionally delimited genes, so care must be applied in delimiting what one counts as the same family of proteins, and hence genes.[28] For example, some definitions require overlap of at least two exons for belonging to the same gene in cases of multiple exon readouts from the same start site (a condition that is violated by some genes that have lots of short exons with complex combinatorics, contributing to medium sized proteins that biochemists consider to belong to the same gene family). Furthermore, as soon as one goes beyond protein encoding genes to try to take account of active sites that include such widely scattered entities as promoters, enhancers, silencers and other regulatory units that need to interact to create some compound proteins and have sometimes been considered to be part of protein encoding genes, one loses contiguity and other similar criteria that were retained by such definitions as the one we just provided. And if one is asking for gene counts, how many regulatory genes are there? There is no stock answer, as there is thanks to the convention that we just cited for explaining how only ~20,000 genes can yield the more than 200,000 proteins in our bodies.

Consequently, genetics education must aim to accommodate effectively and accurately current knowledge, advancements and practices. Perhaps we should move toward a process rather than a material entity account of genes to try to cope with the complex developments that this yields. But it is clear that no neat single definition will work and that authority in developing adequate answers as to what we do and should count as genes is distributed among experts from a variety of different disciplines who ask key questions and are armed with close knowledge of cases in which we hold genes responsible for various outcomes or states of affairs or processes. To get across the excitement of all the material on the forefront AND the need to have command of an enormous range of experimental facts AND the need to bend to shared and distributed authority is a hugely important job, of major importance for education and the public understanding of genetics. To overcome such problems we propose an instrumental concept of genetic material in the next section, a concept that could replace the various gene concepts in substantial parts of our textbooks and in the classroom (see Keller 2010, p. 77, for a similar suggestion for replacing talk of genes with talk of DNA).

---

[28] A recent study (Djebali et al. 2012) found that a shocking 85 % of 492 protein-encoding transcripts for human chromosomes 21 and 22 were chimeric in the sense that they contained transcripts from more than one gene, using the standard boundary definitions for genes. However, as the authors warn, the technical tools involved may yield a significant number of false positives and the proportion of these transcripts that are translated and actually yield proteins is not yet known. These authors, like some others, suggest that the appropriate functional units that should be investigated are *transcripts* rather than *genes*, a stance that would be justified if, as they argue is likely, a substantial proportion of the chimeric transcripts they examined are functional.

7 Towards an Inclusive Concept of Genetic Material to Replace the Concept of Gene in Genetics Education

People learn about genetics in formal (school), informal (science museums), and non-formal (mass media) ways. One of the aims of formal and informal science education is to educate scientifically literate citizens. One can distinguish between two types of science/scientific literacy (Roberts 2007). The first refers to issues within science and it is related to the content of science taught in classrooms. In the case of genetics this should be knowledge about DNA, genes, chromosomes, patterns of inheritance etc. The second is related to questions that students may encounter as citizens, e.g. about the implications of scientific knowledge for society. In the case of genetics, this should be knowledge about e.g. the ethical questions related to genetic testing or to disclosing genomic information about individuals. Thus, future citizens, literate about science, should have a sufficient level of updated and accurate knowledge about the content of science in order to be able to make informed decisions about socio-ethical issues.[29]

For instance, in order to make an informed decision about whether a couple, both of whom are heterozygotes for β-thalassemia, should go through preimplantation genetic diagnosis in order to ensure that their children will not have the disease, they should be aware that they would have to go through an in vitro fertilization procedure and that some healthy and potentially viable embryos might not be eventually transferred to the mother. They should also know that, in case they carry different defective alleles of the β-gene (or, better, of the DNA sequence that is implicated in the production of β-globin peptide chains), their child would be a compound heterozygote who might or might not suffer from the disease. To achieve this, people need to realize the enormous complexity of development, as well as that phenotypes are not simply "controlled" by genes.

However, this is not currently the case. It seems that the contemporary presentation of genetics in schools eventually teaches students that there are genes that "control" or "code for" individual properties. Important phenomena such as epistasis, pleiotropy, plasticity, epigenetics, gene regulation, gene overlap, alternative splicing, antisense reading, etc. (Barnes and Dupré 2008; Stern 2011) are overlooked or at best treated as exceptions. The contemporary presentation of genetics in biology textbooks does not take into account the reality and complexities of development, as a recent study has revealed (Gericke et al. 2012). Most interestingly, in a recent study of teachers' conceptions of genetic determinism in several countries, it was found that even biology teachers may hold strong views of genetic determinism (Castera and Clement 2012). The conclusions from these two studies should alert textbook authors, curriculum developers and science educators about the prevalence of outdated models that enhance mistaken notions of genetic determinism. If these

---

[29] Scientists should, and must, have a role in communicating contemporary knowledge about genetics to the public (see Reydon et al. 2012 for a relevant proposal, which also includes some of the arguments made here).

models are to remain in textbooks and if teachers are not sufficiently familiar with contemporary knowledge of genetics and development, it should be no surprise that people embrace a strong view of genetic determinism (Moore 2008, see also Moore this volume) or that students' writings reveal important misconceptions (Mills et al. 2008; Dougherty 2009).

One possible cause of this problem is the fact that Mendelian genetics is still what most people are taught at school. This is problematic in various ways (see Jamieson and Radick, this volume). Of course, Mendelian genetics still is a valuable heuristic tool and a useful starting point for teaching genetics. Indeed, the description of alleles that control specific characteristics is comprehensible and even middle school students can easily perform simple crosses using so-called Punnett squares. However, if genetics education does not also accommodate recent knowledge about genetics, students will not be able to understand the contemporary issues. With the increasing availability of direct-to-consumer genetic tests for several types of disease, it is important to provide students with the tools to understand what these tests can and cannot reveal. Perhaps the most crucial issue is to help them understand that the intuitive idea of genetic determinism is simply wrong. We hope to have shown not only that the idea of "genes for" is misleading, but also that genes, as such, are not generally distinct units, except when the context is adequately specified (which is seldom the case), with respect to particular phenotypes!

Perhaps the most crucial, neglected component for understanding genetics is that of development. People often do not realize that genes can do nothing outside their cellular contexts and that even evolution proceeds not primarily due to changes in protein coding genes but rather due to changes in regulatory sequences that control the expression of these genes (Stern 2011; Bateson and Gluckman 2011). Genetics education should make clear that the contribution of genes cannot strictly be distinguished from the contribution of their cellular and external environment. Although genes make a partial contribution to a final outcome, they can do nothing on their own. Consequently, only comparisons are possible. To illustrate this, Keller (2010) uses the metaphor of a drummer and his/her drums. There is no point in asking whether the sound produced is more due to the drummer or due to the drums. What would make sense would only be to compare two drummers playing with the same drums, or the same drummer playing with different drums. It is only then that distinguishing between the contributions of the drummer and the drums would make sense. Similarly, distinguishing between the contribution of someone's genes and someone's environment – food, lifestyle, etc. – generally makes sense when comparing genetic differences in persons with highly similar environments, or environmental differences for persons with highly similar genetic makeup.

This is not what one finds even in otherwise excellent textbooks. In a recently published biology textbook (Walpole et al. 2011), the definition of gene given is the following: "A gene is a particular section of a DNA strand that, when transcribed and translated, forms a specific polypeptide" (p. 67). In the glossary of the same book gene is defined as: "a heritable factor that controls a specific characteristic" (p. 586). This is an excellent example of a referentially indefinite gene concept. The two definitions are not entirely consistent with each other. The definition in the

main text of the book is a definite one, which is explicit about the composition (DNA) and the function (forming a polypeptide) of genes. In contrast, the definition in the glossary is an indefinite one that is not explicit about the composition (the factor could be any kind of molecule) or the function (the characteristic is certainly a phenotypic one but it could be either an enzyme or a macroscopic feature such as eye color). The definition in the glossary is thus a less definite one as it does not identify genes with DNA or the synthesis of a particular peptide. As such, it includes both epigenetic and genetic causes of heredity and it ties genes to functions in a different way than the first definition. For example, it would include as a gene a stretch of DNA that makes a regulatory RNA that blocks translation of the message for a key protein, thus regulating the functions of that protein. Since the key to this sort of control is not a protein, the main definition would not acknowledge this sort of gene. Note that an acetylated histone that causes the conformation of a chromosomal region to make certain DNA inaccessible and thus prevents a key gene from being expressed in an embryo, usually considered an epigene or an epigenetic mark on the histone, would count as a gene on the glossary definition. A third definition, set out in a box next to main text, seems to be an attempt to encompass both these definitions, but it rather makes things more complicated: "Gene [is] a heritable factor that controls a specific characteristic, or a section of DNA that codes for the formation of a polypeptide" (p. 68).

How is the gene concept used in the book? Definite descriptions seem to predominate: "Hemophilia is a condition in which the blood of an affected person does not clot normally. It is a sex-linked condition because the genes controlling the production of the blood-clotting protein factor VIII are on the X chromosome." (p. 82). Genes are composed of DNA and control particular characteristics, in this case the production of a protein that is involved in blood clotting. What is worse, the book gives the impression that genes are all powerful. Here is an example: "The fertilised egg of any organism contains all the information needed for developing that single cell into a complex organism consisting of many different types of cell. This information is all within the genes, inherited from the maternal and paternal DNA as fine threads called chromosomes" (p. 16). This definition refers to the robustness of development but is absolutely blind to developmental plasticity. It gives the impression that literally all information about development is included in genes and ignores the fact that information is not, as such, a property of DNA. Information is a kind of relationship between DNA and the translational machinery of the cell as influenced by relationships between cells and by environmental factors.

Another textbook (Sadava et al. 2011) poses similar problems. 'Gene' is defined in the glossary as: "A unit of heredity. Used here as the unit of genetic function which carries the information for a single polypeptide or RNA" (p.G-12). Although the idea of information is identified with the particular unit, the relational character of the gene-as-information is not reflected in the definition of the gene as a unit. Furthermore, the concept of the gene is not identified with DNA or any other molecule. However, this is not the case in the main text of the book where the concept of gene is more definite and actually identified with DNA: "Genes are specific segments of

DNA encoding the information the cell uses to make proteins" (p. 6); "The sequences of DNA that encode specific proteins are transcribed into RNA and are called genes" (p. 64); "Genes are now known to be regions of the DNA molecules in chromosomes. More specifically, a gene is a sequence of DNA that resides at a particular site on a chromosome, called a locus (plural loci). Genes are expressed in the phenotype mostly as proteins with particular functions, such as enzymes" (p. 242). These characterizations do not recognize the context-dependence of the boundaries that are read out to make proteins. Nor do they acknowledge that there are RNA genes or that genes may be located on plasmids and other non-chromosomal molecules. Information is once more not presented as a relation.

We suggest that, given the analyses of the previous sections, these definitions of genes are problematic. Therefore, we propose that genetics education should utilize the wider and more inclusive concept of the genetic material rather than the concept of the "gene for". One could then base the discussion on the evolution of genetic material, the interaction between the genetic material and its intracellular or extracellular context, and the expression of genetic material to produce RNA, proteins or other molecules, and introduce the distinction between genetic and epigenetic inheritance. Biology education ought not focus on DNA and genes and then make a leap to organisms and their phenotypes, overlooking the developmental processes that produce them. There is more to biology than nucleotide sequences, as there is more to language than letter sequences. All cells in an organism contain the same genes (up to mutations acquired during the organism's lifetime[30]), but their expression is differentiated according to their environment and the regulatory apparatus in the cells of the organism. Epistatic and pleiotropic interactions also influence the phenotype. Thus, it is important for biology education to make clear that development is a complex process in which DNA is an important, but not the only, factor.

Based on all the above, we propose that the concept of the gene could be replaced by an instrumental concept of genetic material, explicitly linked to development. The resulting presentation would be more inclusive and more accurate, and could bypass the difficulties raised by the indefinite or functional and definite or structural gene concepts proposed so far as. The proposed concept:

- refers to particular macromolecules (DNA, but also RNA) which are related to the expression and inheritance of traits
- does not refer to particular functions since all functional parts of the genetic material may be implicated in various phenomena and phenotypes

---

[30]In an important footnote, Peter Godfrey-Smith sketches a calculation that shows that after 40 cell divisions any two cells of an adult human would, on average, differ by about 144 point mutations. Most of these mutations would be in non-coding regions and have no phenotypic effects. Nonetheless, it is important to recognize that the casual assumption that most human cells within a human body are genetically identical is simply false (Godfrey-Smith 2009, fn. 9, pp. 82–83). Additionally, within a human body there are about ten times more bacterial and archaeal cells than human cells, some of which are required for proper development (Gilbert and Epel 2009), further undermining simplistic accounts of the genetic uniformity of the cells within our bodies that are essential for our normal functioning.

- does not refer to contiguous DNA sequences because functional units may encompass different parts of the genome

In addition, the proposed concept takes into account the three distinct research programs of Mendelian and molecular genetics, and provides a narrower description of their aims:

1. ***Structure***: The concept of genetic material refers to molecules which are by definition informational. Remember here that 'information' means sequence information and that it is misguiding to locate information in a particular molecule, DNA for instance, except in relation to a given cellular and extracellular context (see Marcos and Arp, this volume). Any molecule with similar informational relationships can be considered to be genetic material, but as far as is now known, only DNA and RNA qualify.
2. ***Function***: Instead of trying to characterize genes in terms of their functions and/or consequences, which was the source of descriptions of "genes for" (e.g., genes for eye color or height, for modifying the action of other genes, for altruism, etc.), we encourage recognition of the multifunctionality of the genetic material, both when one has isolated particular portions of that material and as a whole. This point applies to any kind of information-bearing nucleic acid that directly affects or is implicated at some phenotype at the molecular, cellular or organismal level and provides a convenient way to take into account the contextual dependence of the functions assigned to the nodes of the complex genetic networks that are related to several types of disease.
3. ***Composition and localization***: Instead of seeking to provide precise boundaries for regulatory and protein-encoding genes, we recommend careful examination of the multiple ways in which informational nucleic acids, wherever they are located, relate to other molecules. This procedure reveals the polyfunctionality of the genetic material and the fluidity of its functional boundaries.

To sum up, we propose that the concept of "gene" be used only heuristically in educational books and materials and, for many purposes, that it be replaced by the concept of genetic material which is more inclusive in terms of composition (as it clearly includes RNA as well as DNA), that the localization of genetic material be determined by its sequence-informational function (which takes into account potential multiple effects in multiple contexts) and that the structure of the genetic material be treated as fluid (since whether a molecule, or part of a molecule, is informational or not depends on its interactions and not solely on its molecular structure).

In this sense, we might replace the definition of gene as unit of heredity or a section of DNA which controls a particular polypeptide or phenotypic feature with a definition of genetic material like the following:

> Genetic material: any nucleic acid [composition] in the cell [localization] that interacts with other cellular components and transmits a specific message determining the sequence of other molecules [structure] and thus results in particular, but often quite variable, outcomes inside or outside the cell [function]. These nucleic acids are (usually) reliably copied and maintained from generation to generation, preserving their structure and resulting in the

same functions in similar environments (robustness), though with a range of variation in functions and consequences that depends on cellular and environmental conditions (plasticity). The functions of particular portions of the genetic material may affect or be implicated in cellular processes with local (cellular) or extended (organismal and even environmental) impact; this allows the assignment of fitnesses to particular differences in the genetic material.

Put more simply:

Genetic material: any nucleic acid with the propensity to be inherited and to interact with other cellular components as a source of sequence information, eventually affecting or being implicated in cellular processes with local or extended impact.

This definition is more accurate and inclusive than the typical definitions of the gene. It allows a clear distinction between genetic and epigenetic inheritance, which is not feasible with many standard textbook definitions of the gene. It would free textbooks from referring to "gene(s) for" particular characteristics or diseases but would allow them to refer to particular parts of the genetic material that, in identified contexts, interact with each other and with other cellular components to affect the production of molecular, cellular, or organismic characteristics or to increase the susceptibility of affected individuals to acquire certain traits or diseases.

Let us illustrate why this conception of genetic material is more accurate than the traditional conception of genes. Beta-thalassemia is considered a monogenic disease because various specific mutations at a single region of chromosome 11 affect the production of β-globin molecules. Hemoglobin is produced by the formation of a molecule containing two β-globin and two α-globin molecules. The more defective the β-globin allele, the fewer β-globin molecules are produced or the less well the β-globin molecules trap oxygen when complexed into hemoglobin and consequently the worse the disease is. It is not easy to define the precise functional effects because there are single mutations that can bring about the disease even in heterozygotes, while homozygotes for other mutations have less severe effects. Thus, alleles at the β-hemoglobin locus are evaluated not just by their molecularly specific effects but also by their functional contribution to defects measured by their relevance to health. Familial hypercholesterolemia is also considered a genetic disease, and it is due to a mutation that affects the structure of the LDL-receptor in the liver. However, people who possess the mutation may have milder problems if they follow a proper diet and if they regularly take medication (e.g. statins). In some cases, the problem may totally disappear, so again the mutation is not by itself sufficient to cause anything at the phenotypic level (or if it causes something, it can eventually be reversed). Finally, things are even more complex in cancer which generally is a genetic but not an inherited disease. Somatic mutations, epigenetic changes (sometimes called epimutations in recent literature), and the environment all have crucial influence on most kinds of cancer.

We argue that the concept of the genetic material that affects or is implicated in these situations is more appropriate than any standard concept of the gene as it can be applied in all of these cases and makes appropriate allowance for various degrees of environmental influence (in the wider sense). Our proposal is based on the

importance of cellular, organismal, and environmental influences on the expression (i.e., the expressed or delivered informational content) of the genetic material. Because of the impact of these contextual factors on the information derived from the genetic material, indefinite descriptions of the interplay between the genetic material and the cellular machineries should replace gene concepts. Of course such descriptions can be extremely specific in those cases in which we know what is happening in sufficient detail.

If the aim is to educate scientifically literate citizens, then we should teach non-experts not only what genetics is about but also refrain from enhancing such intuitive conceptions as that of "genes for". Accordingly, we recommend encouraging non-experts to employ indefinite descriptions based on the influences that the genetic material has on (perhaps multiple) characteristics, including, of course, the particular salient characteristics that were formerly used in identifying "genes for" particular traits, while discouraging the genetic determinism that would be reinforced by the idea of "genes for".

References

Barnes, B., and J. Dupré. 2008. *Genomes and what to make of them*. Chicago/London: University of Chicago Press.

Bateson, P., and P. Gluckman. 2011. *Plasticity, robustness, and evolution*. Cambridge: Cambridge University Press.

Benzer, S. 1955. Fine structure of a genetic region in bacteriophage. *Proceedings of the National Academy of Sciences USA* 41: 344–354.

Benzer, S. 1956. Genetic fine structure and its relation to the DNA molecule. *Brookhaven Symposia in Biology* 8: 3–16.

Benzer, S. 1957. The elementary units of heredity. In *A symposium on the chemical basis of heredity*, ed. W.D. McElroy and B. Glass, 70–93. Baltimore: Johns Hopkins University Press.

Beurton, P., R. Falk, and H.-J. Rheinberger (eds.). 2000. *The concept of the gene in development and evolution: Historical and epistemological perspectives*. Cambridge/New York: Cambridge University Press.

Burian, R.M. 1981–1982. Human sociobiology and genetic determinism. *Philosophical Forum* 13: 40–66.

Burian, R.M. 2000. On the internal dynamics of Mendelian genetics. *Comptes rendus de l'Académie des Sciences, Paris. Série III, Sciences de la Vie/Life Sciences* 323(12): 1127–1137.

Burian, R.M. 2005. *The epistemology of development, evolution, and genetics*. New York/Cambridge: Cambridge University Press.

Burian, R. 2013. On gene concepts and teaching genetics – Episodes from classical genetics. *Science & Education* 22(2): 325–344.

Burian, R.M., R.C. Richardson, and W.J. Van der Steen. 1996. Against generality: Meaning in genetics and philosophy. *Studies in History and Philosophy of Science* 27: 1–29.

Carlson, E.A. 1966. *The gene: A critical history*. Philadelphia/London: W. B. Saunders.

Castera, J., and P. Clement. 2012. Teachers' conceptions about genetic determinism of human behaviour: A survey in 23 countries. *Science & Education* (online first article doi:10.1007/s11191-012-9494-0).

Coelho, P.S.R., A.C. Bryan, A. Kumar, G.S. Shadel, and M. Snyder. 2002. A novel mitochondrial protein, Tar1p, is encoded on the antisense strand of the nuclear 25S rDNA. *Genes & Development* 16: 2755–2760.

Crick, F.H.C. 1958. On protein synthesis. *Symposia of the Society for Experimental Biology* 12: The Biological Replication of Macromolecules: 138–163.

Darden, L. 1991. *Theory change in science: Strategies from mendelian genetics*. New York: Oxford University Press.

Darwin, C.R. 1868. *The variation of animals and plants under domestication, 2 vols*. London: John Murray.

de Vries, H. 1910 [1889]. *Intracellular pangenesis*. Chicago: The Open Court Publishing Co. www.esp.org

Dietrich, M.R. 2000. The problem of the gene. *Comptes rendus de l Académie des Sciences. Série III, Sciences de la Vie* 323(12): 1139–1146.

Djebali, S., et al. 2012. Evidence for transcript networks composed of chimeric RNAs in human cells. *PLoS One* 7(1): e28213.

Dougherty, M.J. 2009. Closing the gap: Inverting the genetics curriculum to ensure an informed public. *American Journal of Human Genetics* 85(1): 6–12.

Fogle, T. 2000. The dissolution of protein coding genes in molecular biology. In *The concept of the gene in development and evolution: Historical and epistemological perspectives*, ed. P. Beurton, R. Falk, and H.-J. Rheinberger, 3–25. Cambridge/New York: Cambridge University Press.

Freifelder, D. 1987. *Molecular biology*. Boston: Jones and Bartlett.

Gericke, N., M. Hagberg, V. Carvalho Santos, L.M. Joaquim, and C. El-Hani. 2012. Conceptual variation or incoherence? Textbook discourse on denes in six countries. *Science & Education* (online first article doi:10.1007/s11191-012-9499-8).

Gilbert, S.F. 2000. *Developmental biology*. Sunderland: Sinauer.

Gilbert, S.F., and D. Epel. 2009. *Ecological developmental biology: Integrating epigenetics, medicine, and evolution*. Sunderland: Sinauer Associates.

Godfrey-Smith, P. 2009. *Darwinian populations and natural selection*. Oxford/New York: Oxford University Press.

Goh, K.I., M.E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabasi. 2007. The human disease network. *Proceedings of the National Academy of Sciences USA* 104(21): 8685–8690.

Goodenough, U., and R.P. Levine. 1974. *Genetics*. New York: Holt, Rinehart, and Winston.

Griffiths, P.E., and E.M. Neumann-Held. 1999. The many faces of the gene. *BioScience* 49: 656–662.

Griffiths, P.E., and K.C. Stotz. 2006. Genes in the postgenomic era. *Theoretical Medicine and Bioethics* 27: 499–521.

Jablonka, E., and M.J. Lamb. 2005. *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. Cambridge, MA: MIT Press.

Jablonka, E., and G. Raz. 2009. Transgenerational epigenetic inheritance: Prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly Review of Biology* 84(2): 131–176.

Johannsen, W. 1909. *Elemente der Exakten Erblichkeitslehre*. Jena: G. Fischer.

Kampourakis, K. 2013. Mendel and the path to genetics: Portraying science as a social process. *Science & Education* 22(2): 173–188.

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. 2010. *The mirage of a space between nature and nurture*. Durham: Duke University Press.

Kitcher, P. 1978. Theories, theorists and theoretical change. *Philosophical Review* 87: 519–547.

Kitcher, P. 1982. Genes. *The British Journal for the Philosophy of Science* 33: 337–359.

Mills Shaw, K.R., K. Van Horne, H. Zhang, and J. Boughman. 2008. Essay contest reveals misconceptions of high school students in genetics content. *Genetics* 178(3): 1157–1168.

Moore, D.S. 2008. Espousing interactions and fielding reactions: Addressing laypeople's beliefs about genetic determinism. *Philosophical Psychology* 21(3): 331–348.

Morange, M. 1996. Construction of the developmental gene concept. The crucial years: 1960–1980. *Biologisches Zentralblatt* 115: 132–138.

Morange, M. 2000. Gene function. *Comptes rendus de l'Académie des Sciences. Série III, Sciences de la Vie* 323(12): 1147–1153.

Morange, M. 2001. *The misunderstood gene*. Trans. M. Cobb. Cambridge, MA: Harvard University Press.

Morange, M. 2008. What history tells us XIII. Fifty years of the central dogma. *Journal of Biosciences* 33(2): 171–175.

Morgan, T.H., and C.B. Bridges. 1913. Dilution effects and bicolorism in certain eye colors of Drosophila. *The Journal of Experimental Zoology* 15: 429–466.

Morgan, T.H., A.H. Sturtevant, H.J. Muller, and C.B. Bridges. 1915. *The mechanism of Mendelian heredity*. New York: Henry Holt and Co.

Moss, L. 2001. Deconstructing the gene and reconstructing molecular developmental systems. In *Cycles of contingency: Developmental systems and evolution*, ed. S. Oyama, P.E. Griffith, and R.D. Gray, 85–97. Cambridge, MA: MIT Press.

Moss, L. 2003. *What genes can't do*. Cambridge, MA: MIT.

Neumann-Held, E.M. 1999. The gene is dead – Long live the gene! Conceptualizing genes the constructionist way. In *Sociobiology and bioeconomics: The theory of evolution in biological and economic theory*, ed. P. Kosowski, 105–137. Berlin/New York: Springer.

Neumann-Held, E.M. 2001. Let's talk about genes: The process molecular gene concept and its context. In *Cycles of contingency: Developmental systems and evolution*, ed. S. Oyama, P.E. Griffith, and R.D. Gray, 69–84. Cambridge, MA: MIT Press.

Olby, R.C. 1972. Francis Crick, DNA, and the central dogma. In *The twentieth-century sciences*, ed. G. Holton, 227–280. New York: W. W. Norton.

Olby, R.C. 1975. The protein version of the central dogma. *Genetics* 79(Suppl): 3–14.

Paul, D.B. 1995. Toward a realistic assessment of PKU screening. In *PSA: Proceedings of the biennial meeting of the philosophy of science association, 1994*, vol. 2, ed. A. Fine, M. Forbes, and L. Wessels, 322–328. East Lansing: Philosophy of Science Association.

Portin, P. 2002. Historical development of the concept of the gene. *The Journal of Medicine and Philosophy* 27: 257–286.

Reydon, T.R., K. Kampourakis, and G.P. Patrinos. 2012. Genetics, genomics and society: The responsibilities of scientists for science communication and education. *Personalized Medicine* 9(6): 633–643.

Rheinberger, H.-J. 2000. Gene concepts: Fragments from the perspective of molecular biology. In *The concept of the gene in development and evolution: Historical and epistemological perspectives*, ed. P. Beurton, R. Falk, and H.-J. Rheinberger, 219–239. Cambridge/New York: Cambridge University Press.

Roberts, D.A. 2007. Scientific literacy/science literacy. In *Handbook of research on science education*, ed. S.K. Abell and N.G. Lederman, 729–780. Mahwah: Erlbaum.

Sadava, D., D.M. Hillis, H.C. Heller, and M. Berenbaum. 2011. *Life: The science of biology*. Gordonsville: WH Freeman Publishers.

Sarkar, S. 1998. *Genetics and reductionism*. New York/Cambridge: Cambridge University Press.

Singer, M., and P. Berg. 1991. *Genes and genomes: A changing perspective*. Mill Valley: University Science.

Snyder, M., and M. Gerstein. 2003. Defining genes in the genomics era. *Science* 300: 258–260.

Stadler, L.J. 1954. The gene. *Science* 120: 811–819.

Stern, D.L. 2011. *Evolution, development, and the predictable genome*. Greenwood Village: Roberts & Company Publishers.

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

Strasser, B.J. 2006. A world in one dimension: Linus Pauling, Francis Crick and the central dogma of molecular biology. *History and Philosophy of the Life Sciences* 28: 491–512.

Sturtevant, A.H. 1913a. A third group of linked genes in *Drosophila ampelophila*. *Science* 37: 990–992.

Sturtevant, A.H. 1913b. The linear arrangement of six sex-linked factors in drosophila, as shown by their mode of association. *The Journal of Experimental Zoology* 14: 43–59.

Tycowski, K.T., S. Mei-Di, and J.A. Steltz. 1996. A mammalian gene with introns instead of exons generating stable RNA products. *Nature* 379: 464–466.

Walpole, B., A. Merson-Davies, and L. Dann. 2011. *Biology for the IB diploma coursebook*. Cambridge: Cambridge University Press.

Waters, C.K. 1990. Why the anti-reductionist consensus won't survive: The case of classical Mendelian genetics. In *PSA 1990*, vol. 1, ed. A. Fine, M. Forbes, and L. Wessels, 125–139. East Lansing: Philosophy of Science Association.

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

Waters, C.K. 2000. Molecules made biological. *Revue Internationale de Philosophie* 54: 539–564.

Watson, J.D. 1965. *Molecular biology of the gene*. New York: Benjamin.

Watson, J.D., and F.H.C. Crick. 1953. Molecular structure of nucleic acids. *Nature* 171: 737–738.