Philosophy of Science

June, 1994

GENES MADE MOLECULAR*

C. KENNETH WATERS†‡

Department of Philosophy
and
Minnesota Center for Philosophy of Science
University of Minnesota

This paper investigates what molecular biology has done for our understanding of the gene. I base a new account of the gene concept of classical genetics on the *classical dogma* that gene differences cause phenotypic differences. Although contemporary biologists often think of genes in terms of this concept, molecular biology provides a second way to understand genes. I clarify this second way by articulating a *molecular gene concept*. This concept unifies our understanding of the molecular basis of a wide variety of phenomena, including the phenomena that classical genetics explains in terms of gene differences causing phenotypic differences.

1. Introduction. The impact of advances in genetics during the past few decades is felt in such a broad range of fields that everyone from evolutionist to dairy farmer is talking DNA. Contemporary journals are loaded with articles on genetics describing exciting new findings and molecular-based techniques. But what happened to old-fashioned, classical genetics, the style of investigation and theoretical explanation that formerly dominated the field? Investigators who developed genetics in the first half of this century would not even recognize the routine techniques, basic the-

*Received April 1993; revised June 1993.

†I thank Tom Dahlin for engineering assistance and Kate Beckingham, Lindley Darden, Ron Giere, Bob Herman, David Hull, Harold Kincaid, Pete Magee, David Queller, Joan Strassmann, Mike Simmons, and especially Erich Reck and Tom Wilson for their comments and suggestions on earlier drafts. Early stages of this research was supported by NSF Grant No. Dir 89-12221; later stages were supported by the Graduate School of the University of Minnesota and the McKnight Foundation.

‡Send reprint requests to the author, Minnesota Center for Philosophy of Science, University of Minnesota, Minneapolis, MN 55455, USA.

Philosophy of Science, 61 (1994) pp. 163–185

ory, and molecular-laden terminology presupposed in the contemporary literature. Nevertheless, a number of terms from classical genetics including "gene" continue to play prominent roles in the new molecular style of genetics. The persistence of central terms from the classical theory raises the question of whether these terms are used in the same old ways or in new ways that are aligned with advances in the field. If their usage or meanings have not changed, should we conclude that biologists have failed to bring genetics proper to the molecular level? If their meanings have changed, does it follow that there are two distinct sets of concepts in genetics? If so, are the concepts linked? The aim of this paper is to shed light on what has happened to genetics by examining what has happened to its most central concept, the concept of the gene.

Explaining what has happened to the gene concept will have important implications for two disputes. The first is a disagreement among biologists concerning the question of what is a gene. Despite tremendous strides in genetics and molecular biology, the current literature fails to provide a clear answer. Or, to be blunt, it provides too many ambiguous and conflicting answers. I believe a coherent concept underlies the application of "gene" at the molecular level. My analysis of this underlying concept provides the basis for answering the geneticists' question. The apparent absence of a unifying gene concept at the molecular level has also fueled a controversy among philosophers about the nature of progress in genetics. Philosophers of science once took for granted that a hallmark of scientific progress involved the ability to explain the behavior of entities first conceived at a higher level of analysis (e.g., classical genes) in terms of the behavior of entities conceived at a lower level of analysis (e.g., molecular genes). Such an achievement requires formulating a systematic understanding of the higher level entities in terms of the lower level concepts. Philosophers have argued that this is not achieved in genetics because of a lack of systematic understanding of classical genes at the molecular level (e.g., Hull 1974, Kitcher 1984, Rosenberg 1985, and Dupré 1993; Waters 1990 opposes the antireductionist consensus). My analysis of the gene concepts of classical genetics and molecular biology, however, uncovers a systematic connection between the classical understanding of genes and the contemporary molecular-level one.

One reason philosophers failed to find a connection between the classical and molecular ways of understanding the gene is that biologists have never articulated their molecular-level concept. Another reason is that philosophers have been hampered by a misconception of classical genetics and its gene concept. My study begins with a careful examination of how the explanations of classical genetics worked and provides a new analysis of the underlying gene concept. This is followed by an overview of molecular biology and an analysis of the molecular-level concept of

the gene. My investigation shows that, despite their inconsistent and ambiguous use of "gene", molecular biologists have a concept of the gene which unifies their understanding of a diversity of molecular phenomena. I conclude by discussing the relation between the gene concepts of molecular biology and classical genetics.

2. The Gene Concept of Classical Genetics. Classical genetics was not stagnant. Its concepts, theoretical principles, and methods, which were exemplified in the research and writings of Morgan and his collaborators during the 1910s and 1920s, were extended in subsequent decades by the work of geneticists such as McClintock, Painter, Beadle, and Benzer. (See Carlson 1966 and Darden 1991 for informative accounts of the development of classical genetics.) In this section, I analyze the gene concept that emerged in the 1920s and characterize its main lines of development during the next few decades. (This account is based on Morgan's genetics. See Maienschein 1992 for a broader perspective.)

If my analysis is correct, many present-day accounts read too much into the gene concept of early classical genetics. In order to show that a minimalist conception of the gene was all that classical explanations of inheritance presupposed, classical genetics must be described to a greater extent than would otherwise be required. Hence, section 2.1 illustrates patterns of reasoning in the concrete explanations of classical genetics.

2.1. Classical Genetics. The most fundamental distinction of classical genetics draws a division between the genetic makeup of an organism, often called its "genotype" (more technically termed its "genome") and its manifestation usually called its "phenotype" (more technically its "phenome") (see Lewontin 1992). The relation between the two is causal. The genotype in conjunction with the environment produces the phenotype. According to classical genetics, the genotype of an organism is made up of units, known as genes, which are joined in linear fashion like beads on a string to form one or more linkage groups. Each linkage group associates with a chromosome, and the transmission of genes from parent to progeny can be explained in terms of chromosomal processes during cellular divisions of reproduction. Laws of segregation and independent assortment and principles of genetic recombination and replication are used to explain and predict gene transmission. Classical geneticists explained patterns of inheritance by charting the transmission of genes from generation to generation and attributing the presence of alternative traits to the presence of alternative kinds of genes.

The mode of explanation in classical genetics (as distinguished from its mode of investigation) can be illustrated by examples involving the organism upon which the field developed, the fruit fly *Drosophila me-*

lanogaster. The somatic cells of *Drosophila* have four pairs of chromosomes, named I–IV, and two copies of each gene. Individual copies of each gene pair are located in corresponding positions in the chromosomes of a chromosome pair. The eye color mutant known as purple, for example, is associated with a gene located on chromosome II; that is, two copies of this gene (existing either in mutated or normal "wild-type" form) are located at discrete positions in the two second-chromosomes. The location is termed the "locus" and alternative forms of a gene occurring at a locus are called "alleles". The transmission of genes from parent to offspring is carried out in a special process of cellular division called "meiosis" which produces gamete cells containing half complements of chromosomes (one chromosome from each paired set). The half set of chromosomes from an egg and the half set from a sperm combine during fertilization which gives each offspring a copy of one gene from each gene pair of both parents. Explanations in classical genetics typically relate the presence of genes to the presence of traits in terms of dominant/recessive relations. Purple eye color, for example, is recessive to the wild-type character (red eye color). This means that flies with two copies of the purple allele (the mutant form of the gene, which is designated *pr*) have purple eyes, but "heterozygotes", flies with one copy of the purple allele and one copy of the wild-type allele (designated pr^+), have normal wild-type eyes (as do flies with two copies of the wild-type allele).

To illustrate how classical geneticists explained patterns of inheritance by charting gene transmission and attributing the presence of traits to the presence of genes, consider the cross of a female homozygous for the wild-type *pr* allele (pr^+/pr^+) with a homozygous mutant male (pr/pr). Each offspring receives one copy of chromosome II from each parent. The maternally derived chromosome must contain the wild-type allele (since both second-chromosomes of the mother contain the wild-type allele) and the paternally derived chromosome must contain the purple allele (since both second-chromosomes of the father contain the purple allele). Hence, all offspring are heterozygous (pr/pr^+). Having thus explained the genetic makeup of the progeny, we can draw an inference about phenotypic appearances. Since offspring are all heterozygous (pr/pr^+), and since purple is recessive to wild type, all progeny have the wild-type character. Although this explanatory pattern starts with individuals of known genotype and a mutation known to be recessive to wild type, information about genetic makeup and dominance/recessive relations are originally determined by examining phenotypic patterns revealed by breeding experiments.

C. B. Bridges carried out crosses similar to the one described above, but he kept track of a second mutation as well as the purple one (Bridges and Morgan 1919). The small wing mutant known as vestigial is asso-

ciated with a gene that is also located on chromosome II. The vestigial-wing character is recessive to the wild-type character. In the first experiment, Bridges crossed females that were homozygous for both wild-type alleles (pr^+ vg^+/pr^+ vg^+) with males that were homozygous for both mutant alleles (*pr vg/pr vg*). The offspring were expected to have doubly heterozygous genotypes (pr^+ vg^+/pr *vg*). The presence of "pr^+ vg^+" on the same side of the "/" indicates that these alleles are located on the same chromosome (the one inherited from the female parent). Given the dominance of both wild-type characters, all "**F1**" offspring should exhibit both wild-type phenotypes, which is what Bridges observed. Genes for purple and vestigial characters are said to be "linked" and tend to pass on together through meiosis because they are located close to one another on the same chromosome. Gametes produced by pr^+ vg^+/pr *vg* heterozygotes receive copies of one or the other of the homologous chromosomes. Theoretically, they receive a chromosome with either pr^+ vg^+ or *pr vg*; hence, zygotes produced by the combination of such gametes should not have pr^+ and *vg* (or *pr* vg^+) on the same chromosome. However, in female *Drosophila*, homologous chromosomes frequently swap corresponding sections (containing genes) during meiotic division. Hence, linkage is not absolute. The frequency with which that "crossing over" separates two genes by swapping a segment containing just one of them depends on the distance between them and peaks at fifty percent (because of multiple crossovers). Crossing over played no observable role in Bridges's first cross because the female parents were doubly homozygous.

The effects of crossing over were manifested in Bridges's testcross of **F1** females (pr^+ vg^+/pr *vg*) with males genetically like the purple vestigial parents (*pr vg/pr vg*). Since the doubly heterozygous females used received both of their wild-type alleles from their mothers, these alleles were located on the same chromosome. The two mutant alleles were located on the paternally derived homolog. In the absence of crossing over, half of the gametes produced by the female heterozygote (pr^+ vg^+/pr *vg*) would have copies of both wild-type alleles, and half would have copies of both mutant alleles. Crossing over, however, produced gametes having chromosomes containing one wild-type and one mutant allele (*pr* vg^+ and pr^+ *vg*). Therefore, a portion of the gamete pool produced by doubly heterozygous females is made up of the two kinds of recombinant gametes, (*pr* vg^+ and pr^+ *vg*), in equal numbers. See figure 2.1. When the maternally derived gametes combine with the *pr vg* gametes produced by the males, the result is an **F2** generation with at least half of its members equally divided between the nonrecombinant *pr vg/pr vg* and pr^+ $vg^+/$ *pr vg* genotypes. The remaining portion of the **F2**, which results from recombination, is equally divided between *pr* vg^+/pr *vg* and pr^+ *vg/pr*

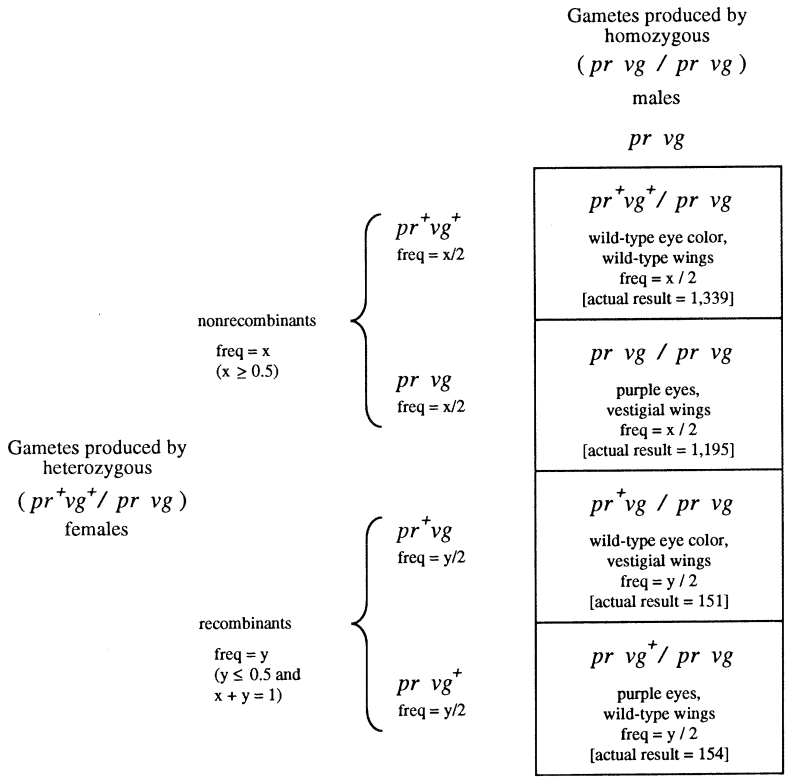

Figure 2.1. Diagram of Bridges's testcross of females doubly heterozygous for purple (*pr*) and vestigial (*vg*) alleles. Information about resulting offspring is given within the boxes. The frequencies given are those expected given classical principles of segregation and recombination (crossing over). Up to 50% of the gamete pool produced by females are recombinants because of crossing over; the ratio of gamete types is one-to-one within the group of nonrecombinant gametes because of segregation (and likewise within the group of recombinant gametes). Bridges's experimental results (Bridges and Morgan 1919) are presented in brackets [ ].

vg genotypes. Having explained the genetic makeup of the **F2** generation, let us now consider phenotypic appearances. At least one quarter have both purple and vestigial characters (those homozygous for both mutant alleles). An equal share have both wild-type characters (those doubly heterozygous). The remaining offspring, which result from genetic recombination, are equally divided between those having the purple eye color and wild-type wing and those having the wild-type eye color and the vestigial wing.

Although I will not go through the details, it is easy to see how Bridges's data could be used to estimate the frequency of recombination, which could then be used to predict and explain results of other breeding experiments involving these genes. By assuming that the frequency of recombination between genes is positively correlated with the distance between them, classical geneticists compared "linkage distances" between genes to map their relative positions. By arranging appropriate breeding experiments, they sought to determine whether genetic mutations producing related phenotypic differences were caused by changes to genes at the same locus. Mutations mapped to the same locus were considered to be mutations of the same gene (i.e., to be alleles), a belief supported by the findings that mutations mapped to the same locus generally exhibit related effects while closely mapped genes did not and the discovery that mutations mapped to the same locus did not complement one another. Although Morgan (1926) did not use the term "complement", he reasoned that if the mutations for white eye and cherry eye were at adjacent loci, crosses between white and cherry would produce daughters with the wild-type character because they would receive a wild-type gene that is dominant to white from the cherry-eyed parent and a wild-type gene that is dominant to cherry from the white-eyed parent. But, as Morgan observed, daughters have intermediately colored eyes (ibid., 92). This is the core reasoning underlying the cis/trans complementation test.

2.2. The Classical Concept of the Gene. The explanations in 2.1 represent classical genetics in its simplest form. Nonetheless, they illustrate the way inheritance of phenotypic *characteristics* can be explained by charting the transmission of genes and relating genotypes to phenotypes. In this subsection, I analyze the classical understanding of the gene standing behind these explanations. I first examine the gene concept that emerged by the 1920s, then briefly describe its main lines of development in later years. My account shows that the patterns of explanation illustrated in 2.1 do not require genes to have any particular internal structure or material makeup. My analysis also shows that in contrast to what the leading philosophical accounts assume, classical explanations do not link genotypes and phenotypes by specifying a gene's contribution to an organism's developing form. I begin by explaining what the classical understanding presupposed about gene structure.

As units passed down from generation to generation, genes were conceived as stable entities, capable of self-replication, located at designated positions in chromosomes. By at least the 1920s, most geneticists thought of genes as having physical structure, and the physical makeup of genes was presumed to provide their stability. Practically nothing, however, was known about the internal structure of the gene until late in the de-

velopment of classical genetics. Speculative passages about gene structure from the 1920s through the 1930s and the metaphor of beads on a string suggest that individual genes were not assumed to have linear structure. Furthermore, diagrams in texts of the 1920s and 1930s do not picture genes as chromosomal segments (i.e., as longitudinal entities lying parallel to the linear structure of chromosomes or linkage maps). The only ideas about structure essential to the explanatory patterns illustrated in section 2.1 are those that enabled geneticists to invoke various chromosomal processes to explain the transmission of genes (processes such as mitosis, meiosis, and the many variations on these themes, e.g., crossing over and nondisjunction). The idea of a linear ordering of genes in chromosomes is essential to these explanations of gene transmission, but this ordering does not imply anything about the structure of the genes themselves. The classical theory made two assumptions about the internal structure of genes: (1) gene structure is relatively stable, and (2) the structure of each gene is replicated before chromosomal division (e.g., Morgan 1926, 27). Muller (1922) pointed out that classical explanations positing spontaneous mutation made an additional assumption: (3) mutations in the structure of a gene are also replicated.

The most subtle aspect of the classical gene concept involves its connection to the genotype/phenotype relation. This relation is not "one gene/one character" because one gene can affect a variety of characters and a single character can be influenced by a number of different genes (genes at different loci). Eye color in *Drosophila*, for instance, is affected by mutations at many different loci; by 1915, Morgan's group could already cite mutations at 25 separate loci that affected eye color (Morgan et al. 1915, 208). Mutations generally affect several characteristics. The white-eye allele (located on the X chromosome), for example, was associated not just with white eyes, but with a colorless sheath of the testes, sluggish behavior, and perhaps a shortened life span as well. The relation between classical gene and phenotype is "many-many", not "one-one". That is, many different genes can affect the same characteristic and many characteristics can be affected by the same gene. Today's accounts of the genotype/phenotype relation of early (pre-1940) classical genetics usually stop here, but classical geneticists offered an abstract causal interpretation of the many-many relation between gene and phenotype.

Morgan (1926) reported that embryology reveals that every organ of the body is the "culmination of a long series of processes" (pp. 305–306) and he presumed that genes act on the steps along the way. If each step in the development of an organ is affected by many genes, he reasoned, then there could be no single gene for the organ. Likewise, if one gene affects steps in the development of more than one organ, then there could

be no single organ associated with a gene. Hence, the many-many relations:

> Suppose, for instance, to take perhaps an extreme case, all the genes are instrumental in producing each organ of the body. This may only mean that they all produce chemical substances essential for the normal course of development. If now one gene is changed so that it produces some substance different from that which it produced before, the end-result may be affected, and if the change affects one organ predominantly it may appear that one gene alone has produced this effect. In a strictly causal sense this is true, but the effect is produced only in conjunction with all the other genes. (Ibid., 306)

Classical geneticists could only speculate about the immediate impact of genes (here, Morgan speculated that they produce chemical substances). This passage and Morgan's discussion of developmental processes suggest that the immediate impact of a gene is separated from characteristics such as eye color by a series of developmental processes that are also influenced by a number of other genes. This means that specifying a gene's contribution to phenotype in terms of characteristics such as eye color would be impossible.

Hull offers a strikingly different view which has been uncritically accepted by subsequent philosophers, "In Mendelian genetics, though some modifications were introduced, the assumption was that the connection [between genes and gross phenotypic characters] was fairly direct. As molecular biology progressed, the extent of the complexity in the production of a single gross character, like eye color in *Drosophila*, was realized" (1974, 29). But molecular biology simply explains the complexity already posited by classical geneticists (see Waters 1990).

Textbooks of classical genetics contain ambiguous statements that have been misinterpreted as suggesting that a given gene is responsible for a particular organ or feature of an organ in an individual. Sturtevant and Beadle wrote, for example, "The element in the X chromosome [of *Drosophila*] that is responsible for the bar eye is the bar *gene*" (1939, 27). I argue that classical genetics provides no basis for saying that the bar mutation occurs in the gene responsible for eye shape (an alternative interpretation will be offered below). The causal connection between gene and phenotype was conceived to be so complex that it would be untenable to think either (a) that only one gene is responsible for eye shape, or (b) that because a mutation in a gene has a prominent effect on eye shape, the main contribution of the unmutated gene is on eye shape. Many accounts of classical genetics emphasize that geneticists did not understand *how* (i.e., by what mechanism) individual genes made their contributions to phenotype. The limitation actually went much further. Geneticists did

not even understand *what* the genes' contributions were. The phenotypic differences associated with mutations only provided an "index" to the genes, not a characterization of their individual contributions. As Morgan explained: "[T]he character that we choose to follow in any case is only the most conspicuous (for purposes of identification) or the most striking or convenient modification that is produced. . . . [T]he particular difference in the germ-plasm is more significant than the character chosen as its index" (1919, 240).

Before about 1940, geneticists lacked the means to specify the genotype/phenotype relation in terms of a gene/organ, gene/character, or gene/product connection because they did not understand what genes were for (i.e., they did not understand the phenotypic contribution of genes or their immediate impact). How, then, was the genotype/phenotype relation cashed out in the concrete explanations? The answer is that gene *differences* within a genetic and environmental context were linked directly to phenotypic *differences*. Even though geneticists could not determine the contribution of a gene to the phenotype, they could observe the phenotypic impact of a mutation to a gene. The difference between wild-type and mutant versions of a gene was identified as the cause of particular differences in phenotype. So, although geneticists could not identify the impact of the gene associated with the bar mutation, the difference between wild-type and bar versions of the gene could be cited as causing the difference in eye shape in the group of flies under study. When Sturtevant and Beadle wrote that the element "responsible for the bar eye is the bar *gene*", they meant that the bar gene is the element that differs in individuals having bar eyes and that this difference is responsible for the difference in observed phenotypes. The basic dogma of classical genetics was that gene differences cause phenotypic differences.

Leading philosophical accounts of the gene concept obscure the fact that classical geneticists did not feign knowledge of what genes were for. (But see Gifford's 1990 insightful analysis of "genetic trait" and Waters 1990 for accounts consistent with the one developed in this section.) Hull, in his influential account of genetics, states, "Numerous phenotypic characters were studied in breeding experiments and the genes supposedly responsible for them inferred . . ." (1974, 16). But what were studied were character differences, not characters, and what explained them were differences in genes, not the genes themselves. The functions of particular genes were not identified by Morgan and his collaborators. Kitcher's (1982) ingenious account of genes is similarly problematic. Kitcher bases his account on the notion of gene complex, which he says "is an aggregate of that chromosomal material whose nature determines the form taken by some phenotypic character" (p. 350). But it was differences in genes within genetic and environmental contexts, not specified genes or complexes of

genes, that were understood to produce observed phenotypic results. What these differences were thought to determine were not general forms such as red eye color mentioned by Kitcher, but specific differences in form occurring within a group of organisms, such as the difference between red and white eye color in a group of experimental *Drosophila*. This misinterpretation of classical genetics has, as will be explained in section 4, blinded philosophers to what molecular biology has done for our understanding of the classical gene.

The gene concept of classical genetics developed in two important ways as biochemists learned about biological pathways and as geneticists worked on organisms with shorter reproductive cycles. The first development was the articulation of the genotype/phenotype relation directly in terms of gene and gene product. This shift is marked by the one-gene/one-enzyme hypothesis attributed to Beadle and Tatum (1941). According to this hypothesis, genes exert an indirect control on life processes by directly producing the enzymes that catalyze the underlying biochemical reactions. Beadle and Tatum's proposal promised to free genes from the obscurity imposed by the biochemical pathways between their immediate impact and observed end results. Previously, geneticists could only talk about the effects of mutations in genes. Now something concrete could be said about the contribution of genes themselves. The second conceptual development involved the idea that genes have a linear structure running parallel to the linear arrangement of genes on a chromosome. This shift is marked by Benzer's (1955) classic paper on fine structure. Like the one-gene/one-enzyme hypothesis, it is logically independent of the DNA model (Ruse 1973, 206), though Benzer was probably influenced by the linear model of DNA being developed at the time. Geneticists working with bacteria and bacterial viruses could measure the frequency of recombination between parts of a gene, which enabled them to map the relative positions of mutations within genes. Instead of conceiving genes as beads on a string, genes were now viewed as linearly structured entities lined up along the chromosome like trains (individually made up of freight cars) lined up along a railway track.

Kitcher (1982) takes a different approach in his analysis of how the gene concept changed. He believes that conceptual change should be understood as change in referential potential (ibid., 347). He assumes that "gene" tokens referred to chromosomal segments and examines various adjustments in referential potential that occurred over time. While I agree that tracing changes in referential potential is a worthy task, I think we overlook an important element of conceptual change if we fail to recognize that geneticists' understanding of the things to which they were referring changed as well. For example, the attribution of a linear, fine structure to genes marks an important conceptual development, not be-

cause it involved a change in referential potential (Morgan used "gene" to refer to linear structures) or a change in the way genes were picked out, but because it involved a modification in beliefs about the properties of genes.

Just as Darwin's *Origin of Species* ([1859] 1964) contained a scarcity of information about the nature of species and their origin, Morgan's *Theory of the Gene* (1926) had little to say about the structure of genes or their individual contributions to phenotype. The genes of Morgan's genetics were entities with unknown structure and unknown effect. Neither the how nor the what of their individual contributions was understood. Genes could be speculatively related to the overall form of an organism, but the connection between genotype and phenotype was spelled out concretely only in terms of the *classical dogma* that gene differences cause phenotypic differences within genetic and environmental contexts of particular populations. This meant that genes could not be identified by their main phenotypic contribution or direct impact, but only by the quirks of mutation. The concept described here was developed by Morgan and his collaborators in the 1910s and 1920s. As explained above, this concept developed over the years. By the time Watson and Crick discovered the structure of DNA in 1953, geneticists were already advancing their understanding of gene structure and the connection between genes and their products.

3. Genes Molecularized. This section offers an account of the concept underlying the use of "gene" in contemporary molecular biology. Later, in section 4, I will discuss the relation between the molecular gene concept investigated here and the classical concept analyzed in the last section. My account of molecular-level genetics begins, as did my account of classical genetics, with the distinction between genotype and phenotype.

3.1. DNA, RNA, and the Central Dogma. The relation between genotype and phenotype is currently understood in terms of nucleotide sequences and the products for which they code. The relevant nucleotide sequences are DNA segments (except in retroviruses whose genetic material is RNA). DNA molecules consist of two chains running parallel to one another. Each chain is made up of a linear sequence of four kinds of nucleotides. The chains of a DNA molecule are complementary because adenine nucleotides in one chain are positioned across from thymine nucleotides in the paired chain and vice versa. Likewise, guanine and cytosine nucleotides are located across from one another. The genetic information is encoded within linear sequences of nucleotides making up the chains. The specific nucleotide sequences of a double-stranded DNA

molecule are preserved during replication because each strand serves as a template for the synthesis of its complement.

The immediate products of gene expression are single-stranded molecules of RNA, which also contain four kinds of nucleotides. The linear sequences of nucleotides in segments of DNA are "transcribed" into complementary sequences of RNA (though in RNA, uracil takes the place of thymine). Some RNA, called mRNA (messenger RNA) travels from the nucleus to the cytoplasm where its linear sequence of nucleotide triplets is "translated" into a linear sequence of amino acids in a polypeptide. Although much of what I will say applies generally, some of the specifics, such as mRNA leaving the nucleus, applies to eukaryotes (e.g., *Drosophila*), but not to prokaryotes (e.g., bacteria) which lack nuclei. Polypeptides, which are simple chains of amino acids, become functional upon taking on conformational structure. Functional polypeptides or "proteins", in the form of catalytic enzymes, regulate cellular and organismic function by turning on and off the biochemical reactions of the developmental and metabolic pathways. Proteins also play regulatory roles in the expression of genetic information and structural roles in the formation of cellular and extracellular tissues. According to the "central dogma" of molecular biology, the connection between genotype and phenotype goes from linear sequences of nucleotides in DNA to linear sequences of nucleotides in RNA to linear sequences of amino acids in polypeptides to the functional conformation of proteins.

While some sequences of DNA nucleotides are ultimately translated into the primary structure of polypeptides, many are not. For example, some stretches of DNA are transcribed into rRNA (ribosomal RNA), which catalyzes the reaction that links amino acids during polypeptide synthesis. Other stretches are transcribed into tRNA (transfer RNA), which oversees polypeptide synthesis by transporting specific amino acids to the appropriate "codon" of the mRNA molecule (the nucleotide triplet that codes for an amino acid or for the start or stop of synthesis). Some sequences within transcribed segments (in eukaryotes) are lost when a portion of RNA, called an "intron", is snipped out and the remaining portions, "the exons", are spliced together and transported to the cytoplasm for translation. Many other sections of DNA are never transcribed. Some of these are known to play regulatory roles. For example, regulatory sections called "promoters" bind RNA polymerase, an enzyme that initiates transcription. Other sections of DNA play regulatory roles by binding specific regulatory proteins that inhibit transcription of adjacent sections of the DNA molecule. Many other sections of DNA (in eukaryotes) seem to play no role at all.

3.2. The One-Gene/One-Polypeptide Conception of the Gene: Exceptions, Ambiguities, and Inconsistencies. "Gene" is often used by molecular biologists to refer to stretches of DNA coding for single polypeptides.

Perhaps the most typical kind of referent is a DNA segment that is transcribed into one kind of mRNA that is subsequently translated into one kind of polypeptide. "Gene" can refer to sequences coding for single polypeptides in more complicated situations as well. For example, some stretches of DNA are transcribed into mRNA molecules made up of segments that are translated into different polypeptides. The sections of DNA corresponding to these mRNA segments are usually identified as separate genes. It is difficult to find generalizations about usage of "gene" that apply to *all* practicing geneticists and molecular biologists. Philosophers talk as if there is a single field of "molecular genetics". But scientists are clustered into linguistic communities corresponding to their organisms of interest and primary level of investigation (traditional genetics, molecular biology, or biochemistry). I believe my generalizations are true with respect to the community that has the most contact with the relevant phenomena. For example, the generalization stated above concerns how "gene" is applied in situations involving polygenic loci, a situation common in prokaryotes and rare in eukaryotes. The generalization is true of molecular biologists studying prokaryotes; whether it applies to eukaryote geneticists is less clear and less significant.

Though geneticists may have "forced nature" in order to protect their idea that genes are nucleotide sequences in DNA coding for single polypeptides, it was nature that forced this idea on geneticists for at least one class of genes. Regulatory genes, which regulate distant genes, were once believed to exert their influence through a mechanism that does not directly involve polypeptide synthesis. Subsequent research, however, has revealed that regulatory genes control the activity of distant genes by coding for regulatory proteins and RNAs. Despite the generality of the one-gene/one-polypeptide rule, exceptions, ambiguities, and inconsistencies have thwarted efforts to use this rule to articulate a rigorous and unifying concept of the gene.

Exceptions to the one-gene/one-polypeptide conception of the gene abound. While "gene" is often applied to segments of DNA specifying polypeptide chains, it is not the case that all genes code for polypeptides. This generalization is violated by the entire class of RNA genes. (Lewin 1990 avoids calling regions coding for nonmessenger RNA "genes", but the vast majority of molecular biologists freely apply "gene" to nonmessenger RNA coding sequences.) RNA genes are transcribed into nonmessenger RNA molecules such as rRNA and tRNA, which play important roles in polypeptide synthesis, but are not themselves translated into polypeptides. Hence, the application of "gene" to stretches of DNA coding for rRNA and tRNA shows that if a unifying concept of gene exists at the molecular level, it is not pegged to the coding of polypeptides.

Another problem with articulating the gene concept in terms of the one-

gene/one-polypeptide rule stems from an ambiguity in the notion of polypeptide coding region. Which nucleotides are part of the functional coding region for a given polypeptide is unclear. Does the region include only nucleotides making up the complementary codons, or does it include all nucleotides "responsible for the production of the polypeptide"? Regulatory sequences and promoters, for example, might be viewed as partially responsible for the production of a polypeptide, but they do not contain nucleotide triplets coding for its amino acids. These regulatory regions, involved with the regulation and initiation of transcription, play no direct role in translation and do not themselves code for polypeptides. Actual usage of terms reveals trends but provides no sure guidelines. In the past, regulatory regions were sometimes treated as parts of genes, as self-contained genes, or as neither. Current usage is a bit less ambiguous. Terms such as "operator gene" have dropped from currency and molecular biologists rarely, if ever, speak or write as if regulatory regions (such as operators) are self-contained genes. Furthermore, today's molecular biologists tend to talk about regulatory regions as being separate from the "genes" they regulate. (Such claims about standard usage are supported by former texts, such as Strickberger 1968; current texts, such as Watson et al. 1987 and Lewin 1990; dictionaries, such as King and Stansfield 1990; and more importantly by current research articles and oral presentations.)

The presence of introns within eukaryotes gives rise to an additional ambiguity of "gene" (as used within the communities of eukaryote geneticists). Introns, segments of DNA that are spliced out during post-transcriptional processing of primary RNA, are sometimes treated as part of the gene, and sometimes not. This ambiguity is becoming more salient as molecular biologists learn more about the importance of post-transcriptional processing.

Post-transcriptional processing, such as the removal of introns, generates conceptual difficulties because genes have traditionally been conceived as functional chromosomal units coding for polypeptides. The chromosomal units, however, do not always coincide with the units of polypeptide coding because post-transcriptional processing of RNA intervenes. Hence, the attempt to refer to "the" unit generates inconsistent usage. The problem is especially acute in situations involving differential processing of primary RNA. Some stretches of DNA are transcribed into primary RNA that is processed in alternative ways, which can result in different mature RNA and different polypeptide products. For example, the troponin T gene of rat muscle generates two forms of troponin T (alpha and beta forms) depending on which four of its five exons are spliced together (Lewin 1990, 487). Geneticists usually speak of the stretch of DNA containing five exons as a single gene, but this unit contains two

different units of polypeptide coding. The ambiguity associated with the inclusion or exclusion of introns might indicate that different kinds of entities are being conflated under the single heading of "gene".

Complications of gene expression and inconsistent usage of "gene" have led many to wonder whether there is a coherent gene concept at the molecular level or just a hodgepodge of different concepts being designated by the same term. In fact, molecular biologists seem to define gene in whatever way suits them at the time, and single texts typically present several conflicting definitions of the term. Some biologists seem to think that working with an ambiguous term is preferable to adopting a precise definition that will only need continual revision as knowledge advances. Their success lends credibility to this tactic. Nevertheless, I will argue that a uniform concept underlies the application of "gene" in molecular biology. This concept provides unity and coherence to molecular biologists' understanding of the molecular complexities behind the phenomena of genetics.

3.3. The Molecular Gene Concept. The fundamental concept underlying the application of "gene" in molecular biology is that of a *gene for a linear sequence in a product at some stage of genetic expression*, hereafter called the "*molecular gene concept*". This concept applies to potentially replicating sequences of nucleotides encoding linear sequences in products of genetic expression. The sequences are usually continuous or discontinuous sequences of DNA nucleotides (although segments of the indirectly replicated RNA in retroviruses also count). Unlike the classical concept, the molecular gene concept specifies what genes are for. Genes are for linear sequences in products of genetic expression. These products come at successive stages and include primary RNA transcripts, processed RNA, and polypeptides. The molecular gene concept is context dependent. It is applied in different investigative or explanatory contexts to pick out sequences in DNA coding for sequences at different stages of genetic expression. Sometimes it is applied to sequences that encode linear sequences of nucleotides in primary transcripts, sometimes to sequences that encode linear sequences in processed RNA, and sometimes to sequences that encode linear sequences of amino acids in polypeptides. Although the gene for a primary transcript is not the same sequence as the split gene for the corresponding processed RNA (produced, for example, by the removal of an intron from the primary transcript), both sequences are uniformly understood by molecular biologists as being genes for a linear sequence in a product at some stage of genetic expression.

The molecular gene concept is not subject to the exceptions plaguing prior attempts to articulate the gene concept in terms of the one-gene/one-polypeptide rule. The exceptions involved RNA genes which are DNA

segments coding for nonmessenger RNA molecules such as rRNA. The linear sequences of nucleotides in RNA genes code for linear sequences in products, though in this case, the encoded sequences are strings of nucleotides in RNA molecules instead of strings of amino acids in polypeptides. So, the designation of RNA genes as genes is in perfect keeping with the underlying concept of the molecular gene. In fact, the molecular concept highlights what RNA and polypeptide genes have in common.

The molecular gene concept is also not subject to the ambiguities associated with attempts to define "gene" in terms of the one-gene/one-polypeptide rule. The most troubling ambiguity stemmed from the fact that genetic expression involves a multistage process in which sequences encoded by DNA and transcribed into primary RNA in the first stage of expression are later spliced out (as in the case of introns) or left untranslated (as in the case of trailers). This ambiguity has given rise to apparent inconsistent usage of "gene" whereby nucleotides coding for introns and trailers are sometimes considered to be part of the gene and sometimes not. The issue of whether they are constituent parts can be resolved by making explicit use of the molecular gene concept. In the case of introns, for example, we need to distinguish between genes for amino acid sequences in polypeptides (or nucleotide sequences in mature RNAs) and genes for nucleotide sequences in primary RNA. An intron is part of the gene for the primary RNA because the intron's linear sequence of nucleotides codes for a complementary sequence in the primary RNA. But the intron is not part of the gene for the eventual polypeptide product because the intron does not code for a linear sequence in the polypeptide. The question of whether an intron is part of "the" gene is not fully specified. The question needs to be spelled out in terms of a particular linear sequence in a product at some stage of genetic expression. Thus, while use of "gene" can involve ambiguity, the ambiguity stems from failing to identify a particular linear sequence or stage of genetic expression, not from an ambiguity intrinsic to the underlying concept. Molecular biologists understand what the term "gene" refers to in concrete situations because the context of discussion implicitly indicates the relevant stage and product of genetic expression (i.e., the context sets *X*, *Y*, and *Z* in the expression "gene for sequence *X* in product *Y* at stage *Z*").

Interpreting "gene" in terms of the molecular gene concept also clarifies the apparently inconsistent usage of "gene" associated with differential post-transcriptional processing. We can see, for example, why DNA segments expressed by two or more polypeptides are sometimes referred to as one gene and sometimes as two or more. The stretch of DNA called the troponin T gene, for instance, is one gene for a primary RNA product, but includes two overlapping genes for polypeptide products alpha and beta. The molecular concept helps us articulate the idea that genes overlap

when they consist of overlapping nucleotide sequences that are genes for different sequences at the same stage of genetic expression.

In addition to explaining the uniform understanding behind the use of "gene" in a variety of contexts, this interpretation imposes constraints that reinforce developing trends in the use of the term. This is why the molecular gene concept is not subject to the other kind of ambiguity discussed in section 3.2, concerning whether regulatory sequences are genes, constituent parts of genes, or neither. As already mentioned, there is a strong trend against referring to regulatory regions as self-contained genes. Traditional geneticists may disapprove of this trend because regulatory regions such as operators play all the roles attributed to classical genes; they replicate, mutate, and segregate in accordance with the laws of classical genetics. Nevertheless, molecular biologists do not consider them to be self-contained genes. This shift in usage of "gene" is explained by my analysis. Traditional geneticists (e.g., Strickberger 1968), who used terms such as "operator gene", were presupposing the classical gene concept (as analyzed in section 2.2) while molecular biologists, who avoid such terms, are now using the molecular gene concept that excludes the idea that operators and other regulatory regions are self-contained genes.

What about the question of whether regulatory sequences are part of the genes they regulate? Although current trends in usage are not as clear about this matter, my analysis implies that operators and other regulatory regions are thought of as "extragenic" when analyzed at the molecular level. In this case, I predict that the weak trend in favor of the corresponding linguistic stricture will strengthen, that future molecular biologists will restrict the referential domain of "gene" to sequences that are transcribed. This curb on the usage of "gene" has important implications; it means that a fair amount of the genetic material influencing the development and functioning of organisms is extragenic. Since changes in these extragenic regions affect outward characteristics of an organism, mutations can occur outside molecular genes. In fact, recent findings in *Drosophila* genetics indicate that many spontaneous mutations studied by Morgan and his contemporaries were caused by transposable element insertions in regulatory regions.

My prediction about the future usage of "gene" in molecular biology is supported by three considerations. First, it is supported by current trends in usage. An examination of the recent literature shows that *molecular-level investigators* have started calling variations in regulatory regions "variations in upstream regulatory sequences" (for example) as opposed to "variations in genes". The second consideration is that the structures of prokaryotic and eukaryotic genomes make it awkward to include regulatory sequences as part of the genes they regulate. In the case of prokaryotic genomes, several "genes" are typically regulated by the same

regulatory sequences. Including regulatory elements as part of the gene would require the complication of specifying overlapping genes which, in this case, could be avoided by restricting the domain of "gene" to coding regions (as prokaryote geneticists already tend to do). In the case of eukaryotic genomes, regulatory regions usually regulate the activity of single coding regions, but the regulatory regions are often scattered a fair distance from the coding regions. In many investigations, the "genes" (coding regions) are often found long before the elements responsible for their regulation are discovered (the initial tools of investigation probe for coding regions, not for regulatory regions). The spatial separation favors a separation in terminology. The third consideration is that as recombinant technologies advance, biologists are becoming more and more accustomed to the idea that "genes" can be removed from their regulatory contexts and placed into new ones. When a coding region is taken out of its original genomic context, placed into a vector, and then deposited into a new regulatory context, the coding region is referred to as one and the same "gene" throughout the process. Technology and genomic structure will favor my prediction that the current trend toward using "gene" to refer only to coding regions (at the molecular level) will become standard in molecular biology.

Whether a sequence of nucleotides counts as a gene is context sensitive. I have already discussed how the use of "gene" is context sensitive because what the term refers to in a given utterance depends on the investigative or explanatory context (which sets *X*, *Y* and *Z* in the expression "gene for sequence *X* in product *Y* at stage *Z*"). Here, I discuss the separate point that whether a sequence of nucleotides is a gene for a particular sequence (in a given product and stage of genetic expression) depends on its genetic context. RNA processing and the nonuniversality of the genetic code render attributions of "gene for sequence *X* in product *Y* at stage *Z*" context sensitive. In the case of RNA processing, whether a split gene codes for the sequence found in a processed RNA partially depends on whether the primary transcript has been appropriately spliced. How a sequence is spliced depends on the sequence of nucleotides within the intron and on sequences distant from the gene. Such phenomena render the ideas of a gene for mature RNA and a gene for polypeptide product context sensitive, though what counts as a gene for primary RNA remains comparatively context free.

Overzealous reductionists might be surprised to learn that the property of being a molecular gene for sequence *X* in product *Y* at stage *Z* is relational and depends on being part of an appropriate system. This context sensitivity is nothing new for genetics. The association of gene differences with phenotypic differences in classical genetics was context sensitive in a similar way. Classical geneticists knew that whether individuals

homozygous for *pr* had purple eyes instead of white depended on genes at many loci, not just the *pr* locus. Reductionists were right in thinking that advances would be made when the phenomena of genetics could be understood and explained at the molecular level. But holists were correct in thinking that advances in our understanding would come from better understanding properties arising from relations within the genetic system and not from understanding just those properties the parts would exhibit in isolation.

My aim has been to uncover a uniform understanding of gene implicit in the thinking of molecular biologists, not to propose a narrow definition or regulate the use of terms. Some readers might want to define "gene". If so, they could define "gene" as any sequence to which the molecular gene concept applies. The molecular gene concept, however, might be used to define "gene" in narrower ways as well. For example, "gene" could be defined as any continuous sequence of nucleotides coding for a primary RNA product. This definition would restrict application of "gene" to units of chromosomal function. Or, alternatively, "gene" might be defined as any continuous or interrupted sequence of nucleotides coding for whatever product comes last in the chain of genetic expression (e.g., polypeptides in the case of the troponin T gene; processed RNA in the case of tRNA genes). This definition would apply to the units closest to classical phenotypic expression. What advantage would come from adopting a narrow definition is unclear, but my account of the molecular gene concept provides a clear conceptual framework for those who wish to do so.

Other readers might think that molecular biologists should dispense with the term "gene" and use in its place terms describing the various kinds of genomic regions, such as operator region, intron, and coding region. In fact, molecular biologists often use these terms. Nevertheless they also use the term "gene" and my aim has been to clarify the understanding underlying their application of this term. Although I am skeptical that molecular biologists will drop the term (or follow the linguistic advice of philosophers), they might stop applying "gene" at the molecular level. If so, the concept I have labeled the "molecular gene concept" would still play the central and unifying role I attribute to it, but it would do so under a different name such as "coding region".

4. The Relation between Classical Genes and Molecular Biology. The molecularization of the gene marks a conceptual development that has resulted in a second way to think about genes. This molecular-level understanding of the gene differs from the classical with respect to both content and domain of application. With respect to content, the molecular gene concept is centered on the idea that genes are for linear sequences

in products whereas the classical concept is centered on the idea that genes are units whose mutations result in phenotypic differences. The domains of application differ because the classical term applies to regulatory regions such as operators whereas the molecular one does not. Perhaps one of the most interesting features of this conceptual development is that biologists continue to use both concepts. When molecular biologists focus on nucleotide sequences, they think of genes in terms of the molecular concept. But at earlier stages of investigation, when they have not gotten close to specifying nucleotide sequences, they tend to think of genes in terms of the rougher-grained classical concept. Furthermore, biologists in many fields still think of genes in the classical way. For example, population geneticists typically use the classical concept because much of evolutionary change is understood in terms of changes in the frequency of gene *differences*.

Although the connection between the classical understanding of the gene and the understanding provided by molecular biology is not exemplified by a one-to-one correspondence between classical and molecular genes, classical genes are understood at the molecular level. The differences by which classical genes were identified are taken to be differences in nucleotide sequences affecting the transcription of molecular genes. In fact, many of the genotypic differences discovered by Morgan and his contemporaries have now been identified at the molecular level. Some of them are located within molecular genes, while others are located in extragenic regions. Nonetheless, *in every case, the differences affect the transcription of molecular genes*. In this way, *the molecular gene concept unifies our molecular-level understanding of the "classical dogma" that gene differences cause phenotypic differences*.

Recent philosophical attempts to explain, or explain away, the classical understanding of the gene from the molecular perspective have erred in part because they are based on the premise that phenotypic forms were classically viewed as fundamental units of development. These accounts mistakenly assume that classical genes were for forms such as eye color or polarity, and then search in vain to discover which molecular-level entity is responsible for such a form (e.g., Rosenberg 1985). Or they proceed from the other direction and point out that discrete units of DNA do not seem to be for any particular classical phenotypic form (e.g., Kitcher 1992). But as my analysis of the classical gene concept shows, the differences in phenotypic form identified by classical geneticists were not viewed as fundamental units of development; they were understood to be phenotypic quirks caused by differences in the real units of heredity, the genes. And, as explained above, molecular biologists can now determine the exact molecular identity of the relevant differences and explain how

in general such differences produce phenotypic differences within a genetic context.

Philosophers have had a field day with molecular biology's alleged failure to provide an explanation of the classical understanding of genes. Some have concluded that the "causal mainstay" of classical genetics has been eliminated (e.g., Churchland 1988). Others have concluded that its essentials have been protected from elimination because its terms cannot be systematically connected to those of molecular biology. According to this view, classical genetics is preserved as an autonomous science aloof from the molecular grasp of reductionism (see Kitcher 1984). But if my analyses of the gene concepts of classical genetics and molecular biology are correct, one of the key assumptions behind these conclusions is false. There is a uniform way of understanding the basic dogma of classical genetics at the molecular level. Differences in classical genes produce differences in phenotypes because they affect the action of molecular genes. So, what really happened to classical genetics? It went molecular.

REFERENCES

Beadle, G. W. and Tatum, E. L. (1941), "Genetic Control of Biochemical Reactions in *Neospora*", *Proceedings of the National Academy of Science 27*: 499–506.

Benzer, S. (1955), "Fine Structure of a Genetic Region in Bacteriophage", *Proceedings of the National Academy of Science 41*: 344–354.

Bridges, C. B. and Morgan, T. H. (1919), "The Second-Chromosome Group of Mutant Characters", *Carnegie Institution of Washington Publication 278*: 123–304.

Carlson, E. A. (1966), *The Gene: A Critical History*. Philadelphia: Saunders.

Churchland, P. S. (1988), *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Reprint. Cambridge, MA: MIT Press.

Darden, L. (1991), *Theory Change in Science: Strategies from Mendelian Genetics*. New York: Oxford University Press.

Darwin, C. ([1859] 1964), *On the Origin of Species: A Facsimile of the First Edition*. Cambridge, MA: Harvard University Press.

Dupré, J. (1993), *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Gifford, F. (1990), "Genetic Traits", *Biology and Philosophy 5*: 327–347.

Hull, D. L. (1974), *The Philosophy of Biological Science*. Englewood Cliffs, NJ: Prentice-Hall.

King, R. C. and Stansfield, W. D. (1990), *A Dictionary of Genetics*. 4th ed. New York: Oxford University Press.

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

———. (1984), "1953 and All That: A Tale of Two Sciences", *The Philosophical Review 93*: 335–373.

———. (1992), "Gene: Current Usages", in E. F. Keller and E. A. Lloyd (eds.), *Keywords in Evolutionary Biology*. Cambridge, MA: Harvard University Press, pp. 128–131.

Lewin, B. (1990), *Genes IV*. Oxford: Oxford University Press.

Lewontin, R. C. (1992), "Genotype and Phenotype", in E. F. Keller and E. A. Lloyd (eds.), *Keywords in Evolutionary Biology*. Cambridge, MA: Harvard University Press, pp. 137–144.

Maienschein, J. (1992), "Gene: Historical Perspectives", in E. F. Keller and E. A. Lloyd (eds.), *Keywords in Evolutionary Biology*. Cambridge, MA: Harvard University Press, pp. 122–127.

Morgan, T. H. (1919), *The Physical Basis of Heredity*. Philadelphia: Lipincott.
———. (1926), *The Theory of the Gene*. New Haven: Yale University Press.
Morgan, T. H.; Sturtevant, A. H.; Muller, H. J.; and Bridges, C. B. (1915), *The Mechanism of Mendelian Heredity*. New York: Holt & Co.
Muller, H. J. (1922), "Variation Due to Change in the Individual Gene", *American Naturalist 56*: 32–50.
Rosenberg, A. (1985), *The Structure of Biological Science*. Cambridge, England: Cambridge University Press.
Ruse, M. (1973), *The Philosophy of Biology*. London: Hutchinson.
Strickberger, M. W. (1968), *Genetics*. New York: Macmillan.
Sturtevant, A. H. and Beadle, G. W. (1939), *An Introduction to Genetics*. Philadelphia: Saunders.
Waters, C. K. (1990), "Why the Anti-Reductionist Consensus Won't Survive the Case of Classical Mendelian Genetics", *PSA 1990*, vol. 1. East Lansing: Philosophy of Science Association, pp. 125–139.
Watson, J. D.; Hopkins, N. H.; Roberts, J. W.; Steitz, J. A.; and Weiner, A. M. (1987), *Molecular Biology of the Gene*. 4th ed. Menlo Park, CA: Benjamin/Cummings Publishing.