The "Mendelian Gene" and the "Molecular Gene": Two Relevant Concepts of Genetic Units

V. Orgogozo*,[1], A.E. Peluffo*, B. Morizot†
*Institut Jacques Monod, UMR 7592, CNRS-Université Paris Diderot, Sorbonne Paris Cité, Paris, France
†Université Aix-Marseille, CNRS UMR 7304, Aix-en-Provence, France
[1]Corresponding author: e-mail address: virginie.orgogozo@normalesup.org

Contents

Abstract

We focus here on two prevalent meanings of the word gene in research articles. On one hand, the gene, named here "molecular gene," is a stretch of DNA that is transcribed and codes for an RNA or a polypeptide with a known or presumed function (as in "gene network"), whose exact spatial delimitation on the chromosome remains a matter of debate, especially in cases with alternative splicing, antisense transcripts, etc. On the other hand, the gene, called here "Mendelian gene," is a segregating genetic unit which is detected through phenotypic differences associated with different alleles at the same locus (as in "gene flow"). We show that the "Mendelian gene" concept is still extensively used today in biology research and is sometimes confused with the "molecular gene." We try here to clarify the distinction between both concepts. Efforts to delineate the beginning and the end of the DNA sequence corresponding to the "Mendelian gene" and the "molecular gene" reveal that both entities do not always match. We argue that both concepts are part of two relevant frameworks for explaining the biological world.

Current Topics in Developmental Biology
ISSN 0070-2153
http://dx.doi.org/10.1016/bs.ctdb.2016.03.002

1. INTRODUCTION

Since the early days, biologists have tried to extract general concepts from their observation of the living forms in order to increase their understanding of the surrounding world. Familiar examples include the concepts of species, ecosystem, symbiosis, or sexual selection. A concept becomes especially relevant when it can account for observations that were so far unexplained. In the history of biology, new discoveries and new theories have often challenged the underlying ideas and definitions behind existing concepts, and the meaning of certain biological concepts has evolved through time.

The concept of "gene" has, since its inception, been a central organizing notion within biology. The word "gene" was introduced by Johannsen (1911) from Hugo de Vries' "pangenes" (de Vries, 1889), themselves derived from Darwin's original, and erroneous, model of blending heredity, "pangenesis" (Darwin, 1868, 1871). According to Johannsen, the gene is "nothing but a very applicable little word" that helps to explain the inheritance of visible characters, and the sum of all genes is called the "genotype" (Johannsen, 1911). Johannsen insisted that "we do not know a genotype but we are able to demonstrate genotypical differences" and therefore that the genotypes are only accessible to the experimenter by comparing phenotypic traits in different organisms. Johannsen thought that what lies in the zygote are "potentialities" to develop a given phenotype and that it is these potentialities which segregate in the form of genes which are inherited (Johannsen, 1911). Looking back at Johannsen's writings, it is not clear whether in his view genes were necessarily connected to a phenotype: it seems theoretically possible to imagine that certain genes were simply transmitted to the progeny without having any phenotypic effect. Today biologists still struggle to find a consensual and generally accepted definition of the "gene." In 2006, 25 scientists of the Sequence Ontology Consortium, which ultimately aims to describe the features of DNA sequences, spent 2 days of long heated discussions to come up with a consensual definition of the gene (see Table 1; Pearson, 2006). More recently, several articles and books dealing with the definition of the term "gene" have been published (for example, Falk, 2010; Gerstein et al., 2007; Griffiths & Stotz, 2013; Pradeu, 2015), showing that the question of "what is a gene?" remains important.

Table 1 Definition of the Terms "Gene", "Allele," and "Locus" According to Several Biological Databases Consortia and Textbooks

Human Genome Nomenclature Organization http://www.genenames.org/about/guidelines#criteria A *gene* is defined as a DNA segment that contributes to phenotype/function. In the absence of demonstrated function a *gene* may be characterized by sequence, transcription or homology.
Sequence Ontology Consortium (Pearson, 2006) A *gene* is a locatable region of genomic sequence, corresponding to a unit of inheritance, which is associated with regulatory regions, transcribed regions and/or other functional sequence regions.
Ensembl Consortium http://www.ensembl.org/info/genome/genebuild/genome_annotation.html An Ensembl *gene* includes any spliced transcripts with overlapping coding sequence, with the exception of manually annotated readthrough genes which are annotated as a separate locus.
Population Genetics Textbook (Hedrick, 2011) *Allele*: Different form of a gene. *Gene*: Unit of inheritance that is transmitted from parents to offspring. *Locus*: Place where a particular gene resides in the genome.
Molecular Biology of the Cell (Alberts et al., 2008) *Allele*: One of several alternative forms of a gene. In a diploid cell each gene will typically have two alleles, occupying the corresponding position (locus) on homologous chromosomes. *Gene*: Region of DNA that is transcribed as a single unit and carries information for a discrete hereditary characteristic, usually corresponding to a single protein or a single RNA.
Genetics and Analysis of Quantitative Traits (Lynch & Walsh, 1998, p. 51) DNA sequences that encode for particular products (proteins and RNAs) are referred to as *genes*, and their chromosomal locations are called *loci*. Most organisms have two copies of each of several chromosomes, in which case they are said to be diploid. Since DNA replication is an imperfect process, mutations arise, and as a consequence the two "copies" of each gene carried by diploid individuals need not be identical. The various forms of a gene are called *alleles*.
Quantitative Genetics (Falconer & Mackay, 1996, pp. 1–2) Suppose for simplicity that we were concerned with a certain autosomal *locus*, A, and that two different *alleles* at this locus, A1 and A2, were present among the individuals. […] Then there would be three possible genotypes, A1A1, A1A2, A2A2 (we are concerned here, as throughout the book, exclusively with diploid organisms.) […] Each A1A1 individual contains two A1 *genes* and each A1A2 contains one A1 *gene*.

Continued

Table 1 Definition of the Terms "Gene", "Allele," and "Locus" According to Several Biological Databases Consortia and Textbooks—cont'd

Genes IX (Lewin, 2006, p. 845 and 852, Glossary) A *gene* is the segment of DNA specifying a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer), as well as intervening sequences (introns) between individual coding segments (exons). An *allele* is one of several alternative forms of a gene occupying a given locus on a chromosome.

When given, we quote the exact definition. When not available, we provide the most representative quote of the authors' definition of gene.

The history of the concept of gene, after Mendel (1866) and Johannsen (1911), has been recounted in several recent publications (see for example, Deutsch, 2012; Gerstein et al., 2007; Keller, 2009; Portin, 2002; Weber, 2005). In brief, classical genetics first considered the gene as an abstract unit of inheritance which explained phenotypic similarities between parents and children. Then, with the advent of molecular biology, genes became segments of DNA which are used as template to make RNA, which can then be used to build proteins, with particular biochemical activities. Soon after, the simple original idea that a gene should be associated with a single transcript was overturned by the discovery of multiple exceptions (alternative splicing, overlapping transcripts on opposite strands, protein-coding genes nested within the intron of another gene, transcription of most chromosomal DNA, etc.), stirring debates about which piece of DNA should be considered as a gene. In this chapter, rather than exploring the evolution of the concept of gene over the years, we focus on the meaning of "gene" at present. We show that many definitions are still employed today by professional biologists and that it is important to try to understand the meaning of the term "gene" in each context to try to avoid confusion and misunderstandings. We argue that all present concepts of genes can be classified into two main categories, the "Mendelian gene" and the "molecular gene." Most writings regarding the different meanings of the term "gene" over the history of biology have presented the "Mendelian gene" as the precursor, now dead, of the "molecular gene" (Deutsch, 2012; Falk, 1984; Griffiths & Stotz, 2013; Weber, 2005). We argue here that the "Mendelian gene" concept is still alive and has not been completely replaced by the "molecular gene" concept. We provide several concrete examples to illustrate that the "Mendelian gene" and the "molecular gene" do not overlap and that both concepts are currently useful to explain different aspects of our biological world.

2. THE "MENDELIAN GENE" AND THE "MOLECULAR GENE"

Following the insight of most authors (Falk, 1984; Gilbert, 2000; Moss, 2003; Pradeu, 2015; Stern, 2000; Weber, 2005), we distinguish two main embodiments for the concept of "gene." On one hand, *a gene is considered as a stretch of DNA that is transcribed and codes for an RNA or a polypeptide with a known or presumed function* (Gerstein et al., 2007; Pearson, 2006). This is what we name here a "molecular gene." To our knowledge, all genome databases consider the "gene" as the "molecular gene" (Table 1). The "molecular gene" leads to the production of RNAs and proteins, which is translated into a phenotype at the level of the organism. The impact of mutations (changes in the nucleotide sequence) in the "molecular gene" is revealed at the level of the gene expression, whether they induce a change in the amount of RNA/protein produced or in the actual sequence that is expressed. This change can then affect the phenotype of interest, but not necessarily. Experimentally, a "molecular gene" is usually revealed by its expression, that is production of an RNA of the corresponding sequence.

On the other hand, a gene is considered as a genetic unit which is transmitted from parents to offspring and which is detected through phenotypic differences associated with different alleles at the same locus. This is what we call here a "Mendelian gene." We note that the "Mendelian gene" is different from what Mendel called "factors" (Mendel, 1866; Olby, 1979). In Mendel's notation, what we call today homozygous diploid individuals were written ***a*** or ***A*** (rather than ***aa*** or ***AA***), whereas heterozygous were written ***Aa***, indicating that Mendel was indeed focused on the phenotypic state which is passed on (Morange, 2016; Olby, 1979). Mendel factors may be seen as elements that combine into specific arrangements, where the two original factors can sometimes fuse into a single one if they are identical.

The "Mendelian gene" can only be revealed and dealt with experimentally if a genotype difference exists and is associated with a phenotype difference. In a previous paper (Orgogozo, Morizot, & Martin, 2015), we distinguished an abstract entity that encompasses both genetic and phenotypic levels that we named "gephe." A "gephe" consists of a phenotypic change (two distinct phenotypic states), its associated variation at a genetic locus (two alleles), and their relationships. For example, resistance to imidazolinone herbicides that inhibit acetolactate synthase (ALS) is

associated with mutations in the ALS gene in *Arabidopsis thaliana* (Sathasivan, Haughn, & Murai, 1991). In 57 other plant species, substitutions in the ALS gene have also been either linked or conclusively shown through functional tests to be responsible for resistance to such herbicides (Baucom, 2016). Here the ALS-resistance gephe is composed of two alleles of the ALS gene, two phenotypic states (resistance and sensitivity to imidazolinone) and the relationship between the genetic change in ALS and the phenotypic difference under consideration. The ALS-resistance gephe is present in over 58 plant species. "Mendelian genes" that are detected through phenotypic differences are part of a "gephe."

A genetic locus can be conceptualized as a position on the genome. However, it is important to mention that it is not strictly speaking a spatial localization, since the number of loci is invariant with the level of ploidy. For instance, a diploid individual will not have its number of loci divided by two in his haploid gametes. Because it can carry alternative alleles, the locus is a genomic position at which segregates genetic variation. A genetic locus thus harbors distinct "Mendelian genes," each associated with various phenotypic states. Noticeably, certain biologists sometimes assume that the "Mendelian gene" concept is synonymous to the concept of locus (A. Martin & M. Rockman, personal communication). Such assimilation may arise when trying to find a spatial localization for the idea of genotype difference that is inherent to the concept of "Mendelian gene," and this is especially apparent in sentences such as "The latter approach was recently used in sunflowers, for example, to identify several flowering-time *genes* that colocalize with flowering-time QTLs" (Olsen & Wendel, 2013) or "we mapped the *gene* to a 45.1-kb region between two markers pcc17 and pcc14 on chromosome 11" (Pei et al., 2012). However, a progeny cannot be said to inherit one locus from his mother and one locus from his father, it is the "Mendelian genes" and not the genetic loci which are inherited. The concept of "Mendelian gene" is therefore closer to the concept of molecular allele than to the one of genetic locus.

Importantly, the physical embodiment of the "Mendelian gene" does not necessarily correspond to a "molecular gene." For example, in yeast the deletion of a telomere, a chromosome extremity which contains no "molecular genes," leads to cell cycle arrest (Sandell & Zakian, 1993) (see also later for other examples).

In summary, for it to be defined and tackled in an operational manner, the "Mendelian gene" requires a phenotype difference associated with a genotype difference, whereas the "molecular gene" requires transcription.

3. CURRENT LITERATURE OFTEN CONFUSES THE "MENDELIAN GENE" AND THE "MOLECULAR GENE" CONCEPTS

Table 2 provides a compilation of several quotes extracted from recent scientific publications which employ the term "gene," and Table 3 lists various usages of the word "gene" in fixed expressions. Both tables show that in certain instances the word "gene" corresponds to the concept of "molecular gene" explained earlier, in others to the concept of "Mendelian gene" and in yet other contexts to an intermingled combination of both concepts.

Because the word gene is often used without specifying whether it is the "molecular" or the "Mendelian" gene, confusion can arise, especially at the crossroads between different fields. One interesting example can be found

Table 2 A Few Examples of Current Usage of the Word "Gene" in Recent Research Papers

Science (Blomen et al., 2015) Many of the *genes* not targeted by our library encode olfactory receptors that are unlikely to be cell-essential.
Nature (Boettiger et al., 2016) These Polycomb-repressed domains harbour *genes* encoding key developmental transcription factors, whose misexpression can have detrimental consequences in differentiated cells.
PLoS Genetics (Raab, Resnick, & Magnuson, 2015) ARID1B and ARID2 participate in wide-spread cooperation to repress hundreds of *genes*.
Scientific Reports (Versluis et al., 2015) There has not yet been sufficient time for the corresponding resistance *genes* to spread into environmental reservoirs.
Nature Reviews Neurology (Hou, Friedrich, Gounot, & Schacherer, 2015) Parkinson Disease is generally considered a multifactorial disorder that arises owing to a combination of *genes* and environmental factors.
PLoS Genetics (Schumer, Cui, Rosenthal, & Andolfatto, 2015) Simulations reveal that hybrid populations rapidly and frequently become isolated from parental species by fixing combinations of *genes* that hinder successful reproduction with parental species.

In the first three lines the word "gene" refers to the "molecular gene" and in the last three to the "Mendelian gene."

Table 3 Various Usages of the Word "Gene" in Fixed Expressions

Where "Gene" Means "Mendelian Gene"	Where "Gene" Means "Molecular Gene"	Where "Gene" Can Mean Both
Defective gene	Foreign gene	Chimeric gene
Dominant gene	Gene cluster	Gene amplification
Gene conversion	Gene expression	Gene manipulation
Gene flow	Gene family	Gene mapping
Gene frequency	Gene network	Gene sequencing
Gene pool	Gene number	Lateral gene transfer
Mutant gene	Gene polymorphism	Pleiotropic gene
Recessive gene		Resistance gene
Selfish gene	Reporter gene	
Susceptibility gene		

Please note that in molecular biology, what biologists mean by a "resistance gene" is a transcriptional unit whose mutation can cause a gain in resistance, in which case the word "gene" corresponds here to the "molecular gene."

on the Cambridge University Science Forum "The Naked Scientist," which denotes a situation often encountered by some of us during scientific discussions between molecular biologists and population geneticists. On the forum, someone wondered: "if as a human I share 98% of my genes with a chimpanzee and 60% of my genes with a banana, how come I only share 50% of my genes with my own daughter?" (http://www.thenakedscientists.com/HTML/questions/question/919/). The paradox occurs here because the first two instances of the term "gene" are used in the molecular sense whereas the last one is the "Mendelian gene." Inconsistencies and flawed reasoning can also occur in more specialized writings. For example, science writer David Dobbs wrote that "For a century, the primary account of evolution has emphasized the gene's role as architect: a gene (or gene variant) creates a trait that either proves advantageous or not, and is thus selected for, changing a species for the better, or not. […] But a number of biologists argue that we need to replace this gene-centric view with one that more heavily emphasizes the role of gene expression—that we need to see the gene less as an architect and more as a member of a collaborative remodeling and maintenance crew." (https://aeon.co/essays/the-selfish-gene-is-a-great-meme-too-bad-it-s-so-wrong). Here the "molecular gene" concept (gene expression) is mistakenly used within the explanatory framework featuring the "Mendelian gene" (the gene is "selected for"), and the gene is inaccurately seen as an entity which can produce a phenotype alone (Keller, 2010). As Steven Pinker blatantly put it: "Part of the blame goes

to molecular biologists, who hijacked the term "gene" for protein-coding sequences, confusing everyone." (https://richarddawkins.net/2013/12/adversarial-journalism-and-the-selfish-gene/).

The confusion between the two concepts is easily noticed in scientific publications and database resources. For example, the population genetics concept of gene flow, that is, "movement of genes among populations due to dispersal processes" (Petit & Excoffier, 2009) implies that the gene here is the "Mendelian gene" since this is what is transmitted from parent to offspring and therefore from one population to another. The "molecular gene" does not flow between populations, but its various copies/alleles can. If gene flow between populations of mosquitoes was to be observed, it would be the dynamics of the presence/absence of the actual sequences (each of them being a specific allele) which would be characterized. If by "gene" one means the "molecular gene," then the term "gene flow" should be replaced by "allele flow."

In model organisms' databases confusion also exists. Consider the *Drosophila melanogaster* gene *white*. On the database Flybase (http://flybase.org/reports/FBgn0039044.html), we can find, among many other features, the sequence of *white*, its position, molecular functions, biological role, and homology with genes in other species. What is meant by the "sequence" of *white* is the sequence of the wild-type (or reference) allele of the gene *white*. On the other hand, the molecular function and the biological role correspond indeed to the "molecular gene" *white*: they were characterized from the analysis of multiple alleles (some of which resulting from mutagenesis of the reference allele) and biochemical activity of different White proteins, all encoded at the *white* locus. When referring to the *white* gene (or any other gene) within the molecular framework, one pictures the "wild-type" sequence (and now, the database entry regarding this gene). Much like species before population thinking (Mayr, 1975), in the strict taxonomical sense, the "molecular gene" appears under the image of a type, or wild-type, sequence deposited into a database with essential properties (or functions). The corresponding alternative versions (alleles) are thought as variations from that reference sequence which share the same essential properties (locus, function, homology). As pointed out by multiple authors regarding the species (Hull, 1965; Sober, 1980), this fits an essentialist, and very Aristotelian picture of natural kinds, which are first envisioned as ideal types narrowly defined. In contrast, the "Mendelian gene" is defined based on an observed variation in phenotype and genotype, thus through a nontypological approach (also called variation approach, or population approach, to

refer to Mayr's dichotomy). Vagueness of definition seems much more tolerated for the "Mendelian gene," which can correspond to any piece of chromosome transmitted from parents to offspring, generally associated with a phenotype.

Another famous example is Dawkins' (1976) "selfish genes." There is no competition in Dawkin's sense between different "molecular genes" within an organism. Indeed, the *white* gene does not compete against the *p53* gene for survival in populations of *D. melanogaster*. It is the different alleles of a "molecular gene" that may compete against each other. Multiple authors have therefore switched to use the selfish allele terminology (Sterelny & Kitcher, 1988). In Dawkins' own words, "when two genes, like the brown eye and the blue eye gene, are rivals for the same slot on a chromosome, they are called alleles of each other." If talking about the "Mendelian gene" then the "selfish gene" terminology is correct. Because each diploid individual has two Mendelian genes at a given locus, competition will occur between them if they are different (meaning there are in different allelic states) and competition will not occur if they are the same. When saying that "one human being inherits 50% of her genes from her father and 50% of her genes from her mother," one is implying that each parental copy should be considered as one "Mendelian gene," even though the maternal copy and the paternal copy might in some cases correspond to the same allele.

In general, evolutionary biologists mean "Mendelian genes" when they speak about "genes," whereas molecular, cell, and developmental biologists mean "molecular genes." The concepts of pleiotropy and epistasis are particularly revealing in this respect. In broad terms, both fields consider that epistasis occurs when the effect of one *gene* on a phenotype is dependent on the presence of another *gene* (Cordell, 2002; Phillips, 2008) and that pleiotropy occurs when one *gene* affects two or more seemingly unrelated phenotypic traits (Paaby & Rockman, 2013; Stern, 2000, 2010). However, in this definition of pleiotropy and epistasis, the term "gene" is used either as the "Mendelian gene" or as the "molecular gene," and this produces radically different concepts. For example, when biochemical geneticists say that the *cid1* gene is epistatic to the *snf1* gene in the yeast *Saccharomyces cerevisiae* (Avery & Wasserman, 1992), what they mean is that first, loss-of-function mutations in these two genes produce distinct phenotypes, and second, the phenotype of the *cid1 snf1* double mutant is similar to the phenotype of the *cid1* gene. In contrast, in population genetics alleles can display epistatic relationships even though they do not correspond to null alleles that fully remove gene activity (Cordell, 2002; Phillips, 2008). To avoid confusion,

one has to be aware that multiple definitions of epistasis and pleiotropy are currently used and that it is important to pay attention to the context to understand what is meant in each case.

Because biology research fields are relatively well-defined and separated, the problem of using the same word for two different meanings does not always arise. However, in certain research areas, the problem is present and acute. In genome-wide association studies, analyses are mostly performed on "Mendelian genes" (Table 2), but results are often interpreted in terms of "molecular genes," with transcriptional units forming an essential part of the concluding explanatory statement that relates the phenotype to the genotype. The problem also occurs in evolutionary biology, especially in evolutionary genetics and eco-evo-devo, which aims to uncover the rules that underlie the interactions between an organism's environment, genes, and development and to incorporate this knowledge into the theory of evolution (Abouheif et al., 2014; Carroll, 2005). Because these fields have a tradition of coupling population genetics, molecular genetics, and developmental biology into one experimental framework, the term "gene" is used to denote either the "Mendelian gene" or the "molecular gene" depending on the context. For example, BMP4 is a "molecular gene" involved in beak shape differences between Darwin's finches species, in the sense that differences in BMP4 expression levels during beak development have been associated with distinct bill shapes, but BMP4 has not been shown to be a "Mendelian gene" involved in beak shape evolution, in the sense that the causing genetic locus and the causing mutation(s) have not been identified (Abzhanov, Protas, Grant, Grant, & Tabin, 2004). It is entirely possible that the change in BMP4 expression levels that is thought to have occurred during beak shape evolution was actually caused by a mutation in another "molecular gene" acting upstream of BMP4. Confusion between both meanings of the term "gene" may also arise in other interdisciplinary fields of biology, such as human genetics. In this chapter, we try to clarify the distinction and the relationship between the "Mendelian gene" and the "molecular gene."

4. HOW MANY GENES, ALLELES, AND LOCI WITHIN A GENOME?

According to most recent estimates, humans are now thought to carry approximately 19,000 genes in their genome (Ezkurdia et al., 2014). In such a statement, genome refers to the nuclear genome and genes to "molecular

genes," or protein-coding sequences. Let us consider one human being. Although his father and his mother gave him 19,000 genes each, we would agree that he has 19,000 genes and not 38,000. If by "gene" we mean the "Mendelian gene," then it is difficult and probably impossible to estimate the number of genes within a human genome, as there is no correlation between the number of "molecular genes" and the number of "Mendelian genes." If by "Mendelian gene" one means any DNA sequence difference, then the number of "Mendelian genes" within a genome is huge and correlated to the level of nucleotide polymorphism within the population. If one means any change in a chromosome region which is associated with a phenotypic change, then the estimation of the number of "Mendelian genes" is extremely difficult, in particular because of the immensity of the phenotype space (Houle, 2010), of G × G interactions and of the various environmental conditions that can affect phenotypes through G × E interactions. If we take one of those "Mendelian genes" and identify it as the one inherited from the father, then there is an equivalent copy which is inherited by the mother. Now, under this view, a diploid organism has in general, at each locus, two "Mendelian genes" which can be identical (homozygous genotype) or different (heterozygous), corresponding to one "molecular gene." To avoid confusion, the total number of genes is often given for the nuclear *haploid* genome.

Compared to the notion of gene, the concept of allele may, at first thought, seem more clearly defined, but it is not certain. According to certain biologists, a diploid homozygous individual carries *one* allele (and thus two copies of the same allele) whereas others affirm that a diploid homozygous individual has *two* alleles (which are identical). A key question which highlights the confusion is "what makes us diploid: the number of genes or the number of alleles?" One possibility is to reply that there are two Mendelian genes and only one allele (considering that an allele represents one version of a gene), and this fits the Mendelian definition. An alternative, close to the molecular view, is to say that there is one "molecular gene" and two copies of the same sequence, that is, two alleles which are identical. At a given locus, the number of molecular alleles is thus equal or higher than the number of Mendelian alleles. In summary, the Mendelian allele refers to an allelic version whereas the molecular allele refers to one of the copies (which can be identical). These two discordant views are found in various biology textbooks (Table 1), showing that there is no consensus.

Similarly, the term "locus" is loosely defined (Table 1). The word "locus" refers to a genomic or genetic position. A locus can be part of a "molecular

gene" or can correspond to several. As stated by the *Rules and Guidelines from the International Committee on Standardized Genetic Nomenclature for Mice*: "A locus is a point in the genome, identified by a marker, which can be mapped by some means. It does not necessarily correspond to a gene; it could, for example, be an anonymous noncoding DNA segment or a cytogenetic feature. A single gene may have several loci within it (each defined by different markers) and these markers may be separated in genetic or physical mapping experiments. In such cases, it is useful to define these different loci, but normally the gene name should be used to designate the gene itself, as this usually will convey the most information." (http://www.informatics.jax.org/mgihome/nomen/gene.shtml). Examination of the concept of "quantitative trait locus" (QTL) also reveals that a locus can encompass several "Mendelian genes." A QTL is a section of chromosome (the locus) that correlates with variation in a quantitative phenotype (Falconer & Mackay, 1996). In cases where one large-effect QTL is later found to be made of several closely linked QTL with smaller effects (McGregor et al., 2007; Orgogozo, Broman, & Stern, 2006), the original locus is found to be made of several "Mendelian genes." In its smallest size, a locus represents one nucleotide position within a genome and in its largest it can be an entire chromosome.

5. "GENES" AS CAUSAL AGENTS OF PHENOTYPES

The "Mendelian gene" and the "molecular gene" concepts are each part of two distinct frameworks for explaining the causes of phenotypes. The "Mendelian gene" explains phenotypic differences between individuals that can interbreed (members of a given population, parents, offspring, etc.) whereas the "molecular gene" explains the existence of a particular phenotype in a given individual (if the gene were to be absent then the phenotype in question would not be as such). Both concepts are part of a causal–mechanistic explanation of the living world (Salmon, 1994, 1997), as opposed to other types of explanations such as the Hempel–Oppenheim deductive-nomological model (Hempel & Oppenheim, 1948). At least two types of causal–mechanistic explanations can be distinguished, the "constitutive" one, which describes the temporal series of successive mechanisms that generate the phenomenon, and the "etiological" one, which identifies factors whose changes modify the phenomenon that needs to be explained (Waters, 2007; Woodward, 2005). In both cases, causes represent pertinent elements that account for the building up of the phenomenon to

be explained. The "molecular gene" is rather involved in a constitutive explanation and the "Mendelian gene" in an etiological explanation. The "Mendelian gene" concept is often used in a framework which does not allow the reconstitution of the entire chain of causal operations linking the genetic level to the phenotypic level. In contrast, the "molecular gene" is part of a continuous series of explanatory processes: the gene is transcribed into mRNA molecules, which are then translated into proteins, and the accumulation of proteins leads to such-and-such effects at the level of the cell and consequently at the level of the organism. Even though certain authors pointed out that current explanations on how molecular genes play a role in elaborating phenotypes are still not as extensive and constitutive as they could effectively be (for example, the effects of cytoplasmic water, gravity, etc., are generally not taken into account) (Gilbert & Epel, 2009; Keller, 2010; Lewontin, 2001; Oyama, 2000), explanations of phenotypic traits involving "molecular genes" are generally more constitutive than those involving "Mendelian genes." Both concepts are important and bring significant insights in their respective fields of research. The "molecular gene" connects better to molecular and cellular processes than the "Mendelian gene," while the "Mendelian gene" connects more directly to the phenotype at the level of the organism than the "molecular gene."

6. SEARCHING FOR THE CONCRETE OBJECTS REPRESENTED BY "GENES"

For any type of concept, the human mind has a tendency to try to make it correspond to a concrete object, that is, an object which can be isolated in time and space by our sensory system. Yet a concept does not necessarily represent such a concrete entity (Cassirer, 1910). For example, the concept of natural selection (Darwin, 1859; Lewontin, 1970) is fully relevant for our understanding of the living world even though it does not represent a concrete object. The concept of "gene" is particularly interesting in this respect. Even though the notion of "gene" was primarily apprehended as an abstract entity that explains the origin of visible characteristics observed in living organisms and how such phenotypic traits are passed from parents to child, biologists have, since the presence of this word in the scientific literature, struggled to find the physical molecular object embodied by the concept of "gene." Today, both the "Mendelian gene" and the "molecular gene" concepts are extremely used and useful to understand the origin of phenotypic traits, in their respective explanatory frameworks, yet biologists

have not found a consensual agreement about the molecular entity that these concepts are supposed to represent, respectively.

The "molecular gene" concept singles out particular pieces of DNA sequences that specify the production of a RNA and possibly a protein. For each newly sequenced species, researchers usually want to address the now standardized question of the number of genes present in its haploid genome (Wade, 2003), and this requires a clear definition of the "molecular gene." Yet problematic issues remain for defining the beginning and the end of a nucleotide region corresponding to a "molecular gene" and for deciding whether a given stretch of DNA can be considered as a gene or not. These difficulties have to do with *cis*-regulatory sequences, overlapping and spliced genes, parasitic and mobile DNA fragments, pseudogenes, non-coding regions with supposedly important function according to their pattern of evolutionary changes across populations and species (Ezkurdia et al., 2014; Gerstein et al., 2007).

The concrete object represented by the "Mendelian gene" is a particular piece of chromosome which, when replaced by another piece, causes a change in phenotype. A survey of the catalog of mutations that have found to be responsible for natural evolutionary changes between species and populations in animals and plants (Martin & Orgogozo, 2013; Stern & Orgogozo, 2008) shows that the "Mendelian gene" can correspond to a single nucleotide, a *cis*-regulatory region, a "molecular gene," a gene cluster (in the "molecular gene" sense, Table 2) or even an entire chromosome (Orgogozo et al., 2015) (and see later).

In certain explanations, the "Mendelian gene" will represent a piece of DNA that is 1 kb long and in others an entire chromosome. Similarly, in the case of overlapping transcripts, the "molecular gene" can be seen by some biologists as a combination of all the overlapping transcribed regions and by others as a single transcribed region, with the other ones being considered as distinct genes. In any case, irrespective of its concrete materialist basis, both concepts remain useful as abstract entities that provide general explanations for the causes of phenotypes.

A large part of genetics research has been, and still is, devoted to the identification of QTLs and DNA sequences that underlie the variations in phenotype observed between individuals of a segregating population. This search for the materialistic substrate of the "Mendelian gene" often ends up with the identification of a "molecular gene." In the following sections, we first examine two exemplary cases of such searches and we then explore the relationship between the materialistic substrate of the "Mendelian genes" and the "molecular genes."

7. THE *WHITE* GENE

Modern genetics started in 1910 with the discovery of white-eyed *D. melanogaster* flies by Thomas Hunt Morgan and his finding that the transmission of the X chromosome correlates with the segregation of the *white* mutation (Morgan, 1910). Following the first report (Green, 1996) of white-eyed flies by Morgan, literally hundreds of other white-eyed mutants were found. For example, the second published catalog of *D. melanogaster* mutants already compiles a list of 27 white-eyed mutants that were identified independently between Mar. 1915 and Apr. 1942 (Bridges, Brehme, et al., 1944). As Lewis recounted (Lewis, 1995), the exact meaning of "gene" was unclear in the forties. While writing their common textbook entitled "Introduction to genetics," the two students of Morgan, Alfred Sturtevant and Georges Beadle, used the term "gene" differently, and they only realized the discrepancy in their thinking once their book was published. To the geneticist Sturtevant, the *white* gene meant a specific *white* mutant (the "Mendelian gene") whereas to Beadle, who was rather a biochemist, it meant the constellation of *white* alleles including the wild-type one (quite close to the present definition of "molecular gene").

It is not until 1981 that the DNA sequence corresponding to the *white* gene was identified, representing the first cloning of a *D. melanogaster* gene (Bingham, Levis, & Rubin, 1981). Using in situ hybridization, Gehring and Paro showed that several fly strains carrying the $white^a$ mutation have a copia transposable element inserted on the X chromosome at the position of the *white* gene (Gehring & Paro, 1980). Tight genetic linkage between w^a and copia was confirmed (Bingham & Judd, 1981) and using the already known DNA sequence of the copia element, the DNA region flanking the copia transposon was identified (Bingham et al., 1981). A 11–15-kb piece of DNA containing the *white* locus was then isolated. When inserted in many different chromosomal locations through P-element-mediated DNA transformation, this DNA fragment was found to rescue the $white^-$ eye-color phenotype (Gehring, Klemenz, Weber, & Kloter, 1984; Hazelrigg, Levis, & Rubin, 1984). We believe that such rescue experiments, for the *white* locus and for the other loci, were crucial in the subtle progressive switch from the concept of "Mendelian gene" to the concept of "molecular gene." Before the rescue experiment, the *white* gene was not fully delimited spatially and could mean either a specific *white* mutant allele (the "Mendelian gene") or the *white* locus itself (with its constellation of

mutant and wild-type alleles, as Beadle intuited). After the rescue experiment, the *white* gene can be seen as a well-defined DNA region, and this region produces the transcript which can rescue the *white*$^{-}$ mutant phenotype. This novel definition of the *white* gene matches the concept of "molecular gene." In the eyes of biologists, mutants can be considered as artifacts (they arose in the laboratory, after all), while the wild-type locus may seem more universal. It is thus possible that while these experiments were ongoing mutant alleles were progressively discredited in favor of the wild-type sequence which show so potent effects and that this might have incited molecular biologists to switch to the "molecular gene" concept.

While the *white* "molecular gene" was being captured in the literal sense, the various parts that make up a gene were being dissected. Rescue tests with smaller DNA pieces delimited the sequence required in *cis* to a 9.9-kb region including about 2 kb upstream and 2 kb downstream of the coding region corresponding to the mature RNA sequence (Levis, Hazelrigg, & Rubin, 1985; Pirrotta, Steller, & Bozzetti, 1985). Among all the *white* alleles that had been characterized, certain were found to affect the coding region, others the introns and yet others *cis*-regulatory regions (O'Hare, Murphy, Levis, & Rubin, 1984; Pirrotta & Bröckl, 1984). All these *white* "Mendelian genes" were grouped together as variants of the same *white* locus because they affect eye color, they do not complement each other and they hardly recombine. In other terms, the *white* locus represents a unit of recombination, a unit of complementation, and a unit of function (eye color) (Weber, 2005). Nevertheless, exceptions were found. Certain *white* alleles were found to recombine (Lewis, 1952; Mackendrick & Pontecorvo, 1952) and others to display partial complementation (Green, 1959; Lewis, 1956). A further categorization of the *white* "Mendelian genes" into distinct types also appeared possible, based on the precise location of the mutations and the exact eye-color phenotype. For example, four mutants named *white spotted* (w^{SP}) have deletions or insertions into the region between 0.9 and 1.3 kb upstream from the transcribed region and all four have a distinctive yellow-brown speckled eye color (Davison, Chapman, Wedeen, & Bingham, 1985; O'Hare, Levis, & Rubin, 1983; Pirrotta & Bröckl, 1984; Zachar & Bingham, 1982). These observations, among others, thus indicated that the concept of gene originating from classical genetics, where the gene should be the unit of recombination, complementation, and function, was too simplistic. The solution which was chosen to classify less ambiguously the eye color Mendelian genes into groups was based on the *white* "molecular gene": "Mendelian genes" were considered as *white* alleles

if they affect the sequence of the *white* locus and if they lead to aberrant production of the White protein.

In the molecular biology field, the shift to the "molecular gene" concept was absolute. As a matter of facts, the majority of molecular biology research papers about the *white* gene use only the "molecular gene" concept since its molecular identification in the mid-1980s.

8. THE *CURLY* GENE

The *Curly*^{1} allele produces flies with curly wings that bend upward when heterozygote and is lethal when homozygote. Since its report (Ward, 1923), *Curly*^{1} has become an important dominant marker for the second chromosome and this allele is now present in over 350 *D. melanogaster* stocks at the Bloomington Drosophila Stock Center. At least 21 *Curly* alleles have been identified: they all map to the 23A4-23B2 region of chromosome 2, are homozygous lethal, and fail to complement each other (http://flybase.org/reports/FBgn0283531.html). Among these 21 *Curly* alleles, only two show a curly wing phenotype when heterozygote, *Curly*^{1} and *Curly*^{K} (Hurd, Liang, & Lehmann, 2015).

The "molecular gene" behind the *Curly*^{1} mutation was identified very recently (Hurd et al., 2015). Because a *duox* loss-of-function mutant failed to complement *Curly* and because *duox* was located at position 23A4-23B2, the "molecular gene" *duox* was suspected to harbor the *Curly* mutation. Conclusive evidence came from a rescue experiment, as for the *white* locus, where ubiquitous expression of the gene *duox* restored viability of *Curly* homozygous individuals. Remarkably, a single nucleotide change was identified in both *Curly*^{1} and *Curly*^{K} in the coding region of *duox*, which results in the conversion of Glycine1505 into a cysteine in *Curly*^{1} and to a serine in *Curly*^{K}. The Glycine1505 residue is extremely conserved from yeasts to humans, suggesting that it has an important role in the activity of the Duox protein, which belongs to a family of transmembrane NADPH oxidases. Importantly, *duox* loss-of-function mutants were found to be homozygous lethal but had no curly wing phenotype. Overexpression of the *Curly*^{K} version of the "molecular gene" *duox* (*duox*^{CyK}), but not of the *duox* wild-type sequence, was found to cause a curly wing phenotype, demonstrating that the change in wing curvature is indeed due to a single nucleotide change. Here the "Mendelian gene" *Curly*^{1}, which is associated with curly wing phenotype and homozygous lethality, can thus be narrowed down to a single nucleotide site on chromosome 2.

It is interesting to note that experimental evidence for the connection between the "Mendelian gene" and the phenotypic change has involved here manipulation of the "molecular gene" *duox*, for both the rescue experiment and the remaking of curly wings in wild-type flies. In theory, with actual sequencing techniques and CRISPR–Cas-9-targeted genome editing, it should be feasible to identify the genetic change underlying a given phenotypic change without dealing with the "molecular gene" that is affected by the mutation. Nevertheless, most current studies that aim to identify the sequence change responsible for a given phenotype (ie, the concrete nucleotide sequence of the "Mendelian gene") do manipulate the "molecular gene" that they suspect to be involved, because such manipulations are easier and faster that genome editing at the precise position of the suspected genetic change. In any case, once a mutation has been identified as responsible for a given phenotypic change, the concluding explanation that connects the genetic change to the phenotype almost always involves the transcriptional unit itself, that is, the "molecular gene." In other words, even though it is now possible to delineate the spatial localization of "Mendelian genes" without manipulating the "molecular gene," the concept of "molecular gene" remains nevertheless incorporated into the final explanation that links genotypes to phenotypes.

9. THE MOLECULAR DELIMITATIONS OF THE "MENDELIAN GENE" AND THE "MOLECULAR GENE" DO NOT ALWAYS MATCH

In the case of *white*, all the "Mendelian genes" affecting eye color at the *white* locus correspond to mutations that affect coding regions, *cis*-regulatory regions, and/or introns of the *white* "molecular gene." There is thus a good overlap between both gene concepts: the chromosomal location of the various *white* "Mendelian genes" is the *white* "molecular gene," and a mutation that affects the *white* "molecular gene" will make a *white* "Mendelian gene." In the case of curly wings, so far only two "Mendelian genes" at position 23A4-23B2 have been identified and both affect the same nucleotide position (Hurd et al., 2015) (see earlier). In the absence of other mutations causing curly wings at this genomic position, we can hypothesize that only mutations at this nucleotide site will generate curly wings. If this is the case, then the molecular location of the *Curly* "Mendelian gene" is a specific nucleotide position within the *duox* "molecular gene," and thus the "Mendelian gene" and "molecular gene" do not map to the exact same

genomic region. In contrast, if we consider the "Mendelian genes" associated with lethality at the homozygous state at the *duox* locus, then these "Mendelian genes" do map to the same region as the *duox* "molecular gene."

Table 4 lists multiple cases where the molecular delimitations of the "Mendelian gene" and the "molecular gene" do not overlap. In several cases, the "Mendelian gene" has been found to affect more than one "molecular gene." For example, the Williams–Beuren syndrome, which is associated with characteristic facial dysmorphism, cardiac malformation, and a specific behavioral and cognitive profile, is due to a deletion of

Table 4 Various Cases Where the Molecular Delimitations of the "Mendelian Gene" and the "Molecular Gene" Do Not Overlap

Mutation Category	Examples and References
Deletion of several genes	Williams–Beuren syndrome (for more examples, see Table 1 of Weischenfeldt, Symmons, Spitz, & Korbel, 2013)
Insertion of an extra DNA piece that contains several genes	Carotenoid synthesis genes (Altincicek, Kovacs, & Gerardo, 2012; Cobbs, Heath, Stireman, & Abbot, 2013; Moran & Jarvik, 2010)
Extra chromosome	Down syndrome/trisomy 21
Inversion or translocation that leads to the fusion of the coding sequences of two distinct "molecular genes" and the production of a chimeric gene	Philadelphia translocation, which gives rise to the *BCR–ABL1* fusion protein *Trim5–CypA* chimeric gene (Stoye & Yap, 2008)
Inversion or translocation that leads to reshuffling of *cis*-regulatory sequences and coding sequences of several "molecular genes"	Rose-comb mutation (Imsland et al., 2012) *ladybird-C15* inversion (Cande, Chopra, & Levine, 2009)
A single mutation in a *cis*-regulatory element that regulates the expression of multiple "molecular genes"	H element controlling the expression of several odorant receptor genes (Fuss, Omura, & Mombaerts, 2007)
A single mutation at a precise nucleotide position within a "molecular gene"	*Curly*^{1} (Hurd et al., 2015)
Deletion of a centromere DNA element, leading to mitosis delay	CDEII delta 31 (Spencer & Hieter, 1992)
Elimination of a telomere, leading to cell cycle arrest	Yeast telomere elimination (Sandell & Zakian, 1993)

1.5–1.8 Mb on chromosome 7. This deletion affects multiple genes including *ELN* and *LIMK*, and it has been shown that these two genes contribute to the complex phenotype of the Williams–Beuren syndrome (Tassabehji, 2003). Rose-comb is a 7.4-Mb inversion on chromosome 7 in chicken that alters at least two genes: it disrupts the *CCDC108* gene located at one of the inversion breakpoints and it relocalizes the MNR2 homeodomain protein gene, leading to transient ectopic expression of *MNR2* during comb development (Imsland et al., 2012). In other cases, the "Mendelian gene" does not affect any "molecular gene," but simply a DNA sequence that is not a transcriptional unit and whose mutation produces a phenotypic effect (origin of replication, telomere, centromere). For example, a 31-base-pair deletion within centromere DNA element II (CDEII delta 31) of the yeast *S. cerevisiae* causes a dramatic delay in cell division (Spencer & Hieter, 1992).

In summary, efforts to delineate the beginning and the end of the DNA sequence corresponding to a "Mendelian gene" often end up in a genetic unit which corresponds to a "molecular gene." However, this is not always the case. In certain instances, the "Mendelian gene" involves a genetic change in multiple "molecular genes" and in others nucleotide regions devoid of "molecular genes."

10. CONCLUSION

In general, an explanatory framework cannot be reduced to a single concept; it always consists of several concepts and their associated relationships (David & Samadi, 2011). A given concept thus brings explanations mainly through its relationships with other concepts and through a particular way of categorizing the world. The current state of biology research is one where both concepts of genes, the Mendelian and the molecular, continue to be used as explanatory frameworks. Although molecular genetics has brought a much more detailed understanding of what a gene is than classical genetics, the fact that we continue to refer to a premolecular biology framework when talking about genes, especially in population genetics, is an additional proof that the science of genetics, especially its language, is not fully reducible to molecular genetics or genomics (Brigandt & Love, 2008; Sarkar, 1998). Importantly, fields such as population genetics and evolutionary biology which have tried to understand how genes are selected and segregate within a population seem to be more attached to a genetic tradition which employs the "Mendelian gene" as a central concept, whereas fields such as molecular genetics and developmental biology which are focused

on the question of the function of the genes and their general role within a genotype–phenotype relationship have been more focused on the "molecular gene." Despite the difference that we highlight here between these two relevant concepts of genetic unit, it is almost surprising that most of us continue to exchange ideas and communicate our work without too much difficulties regarding what we mean when we talk about genes. Yet asking our colleagues about the number of genes and alleles at one locus in a homozygous diploid seems enough to trigger confusion. What better proof that both views of the "gene" are still alive?

ACKNOWLEDGMENTS

We thank Matt Rockman, Arnaud Martin, François Graner, David L. Stern, and Thomas Pradeu for comments on the chapter. Part of this work was presented at the ISHPSSB Montréal meeting in July 2015. The research leading to this chapter has received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007–2013 Grant Agreement No. 337579), the John Templeton Foundation (Grant No. 43903), and from the "Who Am I?" Laboratory of Excellence (ANR-11-LABX-0071) funded by the French government through its "Investments for the Future" Program operated by the French National Research Agency (ANR) under Grant No. ANR-11-IDEX-0005-02.

REFERENCES

Abouheif, E., Favé, M.-J., Ibarrarán-Viniegra, A. S., Lesoway, M. P., Rafiqi, A. M., & Rajakumar, R. (2014). Eco-evo-devo: The time has come. In C. R. Landry & N. Aubin-Horth (Eds.), *Ecological genomics* (pp. 107–125). Berlin: Springer.

Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R., & Tabin, C. J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. *Science*, *305*, 1462–1465.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter Peter, P. (2008). *Molecular biology of the cell* (5th ed.). New York: Garland Science, Taylor and Francis Group.

Altincicek, B., Kovacs, J. L., & Gerardo, N. M. (2012). Horizontally transferred fungal carotenoid genes in the two-spotted spider mite Tetranychus urticae. *Biology Letters*, *8*, 253–257. http://dx.doi.org/10.1098/rsbl.2011.0704.

Avery, L., & Wasserman, S. (1992). Ordering gene function: The interpretation of epistasis in regulatory hierarchies. *Trends in Genetics*, *8*(9), 312–316.

Baucom, R. S. (2016). The remarkable repeated evolution of herbicide resistance. *American Journal of Botany*, *103*, 181–183. http://dx.doi.org/10.3732/ajb.1500510.

Bingham, P. M., & Judd, B. H. (1981). A copy of the copia transposable element is very tightly linked to the Wa allele at the white locus of D. melanogaster. *Cell*, *25*, 705–711.

Bingham, P. M., Levis, R., & Rubin, G. M. (1981). Cloning of DNA sequences from the white locus of D. melanogaster by a novel and general method. *Cell*, *25*, 693–704.

Blomen, V. A., Májek, P., Jae, L. T., Bigenzahn, J. W., Nieuwenhuis, J., Staring, J., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science*, *350*, 1092–1096. http://dx.doi.org/10.1126/science.aac7557.

Boettiger, A. N., Bintu, B., Moffitt, J. R., Wang, S., Beliveau, B. J., Fudenberg, G., et al. (2016). Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, *529*, 418–422. http://dx.doi.org/10.1038/nature16496.

Bridges, C., Brehme, K., et al. (1944). *The mutants of Drosophila melanogaster*. With foreword by TH Morgan, Washington, DC: Publ. Carnegie Inst. Wash.

Brigandt, I., & Love, A. (2008). *Reductionism in biology*. *In E. N. Zalta (Ed.), The Stanford encyclopedia of philosophy*. Stanford: Stanford University Spring 2013 Ed, *http://plato.stanford.edu/entries/reduction-biology/*.

Cande, J. D., Chopra, V. S., & Levine, M. (2009). Evolving enhancer-promoter interactions within the tinman complex of the flour beetle, Tribolium castaneum. *Development (Cambridge, England)*, *136*, 3153–3160. http://dx.doi.org/10.1242/dev.038034.

Carroll, S. B. (2005). Evolution at two levels: On genes and form. *PLoS Biology*, *3*, e245. http://dx.doi.org/10.1371/journal.pbio.0030245.

Cassirer, E. (1910). *Substance and function* (2010 ed.). Paris: Nabu Press.

Cobbs, C., Heath, J., Stireman, J. O., 3rd., & Abbot, P. (2013). Carotenoids in unexpected places: Gall midges, lateral gene transfer, and carotenoid biosynthesis in animals. *Molecular Phylogenetics and Evolution*, *68*, 221–228. http://dx.doi.org/10.1016/j.ympev.2013.03.012.

Cordell, H. J. (2002). Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics*, *11*, 2463–2468. http://dx.doi.org/10.1093/hmg/11.20.2463.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.

Darwin, C. (1868). *The variation of animals and plants under domestication*. London: John Murray.

Darwin, C. (1871). Pangenesis. *Nature*, *3*, 502–503.

David, P., & Samadi, S. (2011). *La théorie de l'évolution: Une logique pour la biologie*. Paris: Flammarion.

Davison, D., Chapman, C. H., Wedeen, C., & Bingham, P. M. (1985). Genetic and physical studies of a portion of the white locus participating in transcriptional regulation and in synapsis-dependent interactions in Drosophila adult tissues. *Genetics*, *110*(3), 479–494.

Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.

de Vries, H. (1889). *Intracellular pangenesis*. Chicago: The Open Court Publishing Co.

Deutsch, J. (2012). *Le Gène: Un concept en évolution*. Paris: Seuil.

Ezkurdia, I., Juan, D., Rodriguez, J. M., Frankish, A., Diekhans, M., Harrow, J., et al. (2014). Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Human Molecular Genetics*, *23*, 5866–5878. http://dx.doi.org/10.1093/hmg/ddu309.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Harlow, UK: Pearson Education.

Falk, R. (1984). The gene in search of an identity. *Human Genetics*, *68*, 195–204.

Falk, R. (2010). What is a gene?—Revisited. *Studies in History and Philosophy of Science. Part C, Studies in History and Philosophy of Biological and Biomedical Sciences*, *41*, 396–406.

Fuss, S. H., Omura, M., & Mombaerts, P. (2007). Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell*, *130*, 373–384. http://dx.doi.org/10.1016/j.cell.2007.06.023.

Gehring, W. J., Klemenz, R., Weber, U., & Kloter, U. (1984). Functional analysis of the white gene of Drosophila by P-factor-mediated transformation. *The EMBO Journal*, *3*, 2077–2085.

Gehring, W. J., & Paro, R. (1980). Isolation of a hybrid plasmid with homologous sequences to a transposing element of Drosophila melanogaster. *Cell*, *19*, 897–904.

Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., et al. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Research*, *17*, 669–681. http://dx.doi.org/10.1101/gr.6339607.

Gilbert, S. F. (2000). Genes classical and genes developmental: The different uses of genes in evolutionary syntheses. In P. Buerton, R. Falk, & H.-J. Rheinberger (Eds.), *The concept of*

the gene in development and evolution (pp. 178–192). Cambridge: Cambridge University Press.
Gilbert, S. F., & Epel, D. (2009). *Ecological developmental biology: Integrating epigenetics, medicine, and evolution*. Sunderland, MA: Sinauer Associates.
Green, M. M. (1959). Putative non-reciprocal crossing over in Drosophila melanogaster. *Molecular and General Genetics MGG*, *90*, 375–384.
Green, M. M. (1996). The "genesis of the white-eyed mutant" in Drosophila melanogaster: A reappraisal. *Genetics*, *142*, 329.
Griffiths, P., & Stotz, K. (2013). *Genetics and philosophy: An introduction*. Cambridge, UK; New York: Cambridge University Press.
Hazelrigg, T., Levis, R., & Rubin, G. M. (1984). Transformation of white locus DNA in Drosophila: Dosage compensation, zeste interaction, and position effects. *Cell*, *36*, 469–481.
Hedrick, P. W. (2011). *Genetics of populations* (4th ed.). Sudbury, MA: Jones & Bartlett Learning.
Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy in Science*, *15*, 135–175.
Hou, J., Friedrich, A., Gounot, J.-S., & Schacherer, J. (2015). Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. *Nature Communications*, *6*, 7214.
Houle, D. (2010). Colloquium papers: Numbering the hairs on our heads: The shared challenge and promise of phenomics. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(Suppl. 1), 1793–1799. http://dx.doi.org/10.1073/pnas.0906195106.
Hull, D. L. (1965). The effect of essentialism on taxonomy—Two thousand years of stasis (I). *The British Journal for the Philosophy of Science*, *15*, 314–326.
Hurd, T. R., Liang, F.-X., & Lehmann, R. (2015). Curly encodes dual oxidase, which acts with heme peroxidase Curly Su to shape the adult Drosophila wing. *PLoS Genetics*, *11*, e1005625.
Imsland, F., Feng, C., Boije, H., Bed'hom, B., Fillon, V., Dorshorst, B., et al. (2012). The Rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genetics*, *8*, e1002775. http://dx.doi.org/10.1371/journal.pgen.1002775.
Johannsen, W. (1911). The genotype conception of heredity. *The American Naturalist*, *45*, 129–159.
Keller, E. F. (2009). *The century of the gene*. Cambridge, MA: Harvard University Press.
Keller, E. F. (2010). *The mirage of a space between nature and nurture*. Durham and London: Duke University Press.
Levis, R., Hazelrigg, T., & Rubin, G. M. (1985). Separable cis-acting control elements for expression of the white gene of Drosophila. *The EMBO Journal*, *4*, 3489–3499.
Lewin, B. (2006). *Genes IX*. Sudbury, MA: Jones & Bartlett.
Lewis, E. B. (1952). The pseudoallelism of white and apricot in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, *38*, 953–961.
Lewis, E. B. (1956). An unstable gene in Drosophila melanogaster. *Genetics*, *41*, 651.
Lewis, E. B. (1995). Remembering Sturtevant. *Genetics*, *141*, 1227.
Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, *1*, 1–18.
Lewontin, R. C. (2001). *The triple helix: Gene, organism, and environment*. Cambridge, MA: Harvard University Press.
Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sunderland, MA: Sinauer Associates.
Mackendrick, M. E., & Pontecorvo, G. (1952). Crossing-over between alleles at the w locus in Drosophila melanogaster. *Experientia*, *8*, 390–391.

Martin, A., & Orgogozo, V. (2013). The loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. *Evolution: International Journal of Organic Evolution*, *67*, 1235–1250. http://dx.doi.org/10.1111/evo.12081.

Mayr, E. (1975). *Evolution and the diversity of life*. Cambridge, MA: Harvard University Press.

McGregor, A., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D., Payre, F., et al. (2007). Morphological evolution through multiple cis-regulatory mutations at a single. *Nature*, *448*, 587–590.

Mendel, G. (1866). Versuche über Pflanzenhybriden. *Verhandlungen Naturforschenden Vereines Brunn*, *4*, 3–47.

Moran, N. A., & Jarvik, T. (2010). Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, *328*, 624–627. http://dx.doi.org/10.1126/science.1187113.

Morange, M. (2016). *Une histoire de la biologie*. Paris: Points.

Morgan, T. (1910). Sex limited inheritance in Drosophila. *Science*, *32*, 120–122.

Moss, L. (2003). One, two (too?), many genes? *The Quarterly Review of Biology*, *78*, 57–67. http://dx.doi.org/10.1086/367581.

O'Hare, K., Levis, R., & Rubin, G. M. (1983). Transcription of the white locus in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(22), 6917–6921.

O'Hare, K., Murphy, C., Levis, R., & Rubin, G. M. (1984). DNA sequence of the white locus of Drosophila melanogaster. *Journal of Molecular Biology*, *180*(3), 437–455.

Olby, R. (1979). Mendel no Mendelian? *History of Science*, *17*(1), 53–72.

Olsen, K. M., & Wendel, J. F. (2013). A bountiful harvest: Genomic insights into crop domestication phenotypes. *Annual Review of Plant Biology*, *64*, 47–70.

Orgogozo, V., Broman, K. W., & Stern, D. L. (2006). High-resolution quantitative trait locus mapping reveals sign epistasis controlling ovariole number between two Drosophila species. *Genetics*, *173*, 197–205.

Orgogozo, V., Morizot, B., & Martin, A. (2015). The differential view of genotype–phenotype relationships. *Frontiers in Genetics*, *6*, 179. http://dx.doi.org/10.3389/fgene.2015.00179.

Oyama, S. (2000). *The ontogeny of information: Developmental systems and evolution*. Durham, London: Duke University Press.

Paaby, A. B., & Rockman, M. V. (2013). The many faces of pleiotropy. *Trends in Genetics*, *29*, 66–73. http://dx.doi.org/10.1016/j.tig.2012.10.010.

Pearson, H. (2006). Genetics: What is a gene? *Nature*, *441*, 398–401.

Pei, C., Wang, H., Zhang, J., Wang, Y., Francis, D. M., & Yang, W. (2012). Fine mapping and analysis of a candidate gene in tomato accession PI128216 conferring hypersensitive resistance to bacterial spot race T3. *Theoretical and Applied Genetics*, *124*(3), 533–542.

Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology & Evolution*, *24*(7), 386–393.

Phillips, P. C. (2008). Epistasis: The essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews. Genetics*, *9*, 855–867. http://dx.doi.org/10.1038/nrg2452.

Pirrotta, V., & Bröckl, C. (1984). Transcription of the Drosophila white locus and some of its mutants. *The EMBO Journal*, *3*(3), 563.

Pirrotta, V., Steller, H., & Bozzetti, M. P. (1985). Multiple upstream regulatory elements control the expression of the Drosophila white gene. *The EMBO Journal*, *4*, 3501–3508.

Portin, P. (2002). Historical development of the concept of the gene. *The Journal of Medicine and Philosophy*, *27*, 257–286. http://dx.doi.org/10.1076/jmep.27.3.257.2980.

Pradeu, T. (2015). Toolbox murders: Putting genes in their epigenetic and ecological contexts. *Biology and Philosophy*, *31*, 125–142.

Raab, J. R., Resnick, S., & Magnuson, T. (2015). Genome-wide transcriptional regulation mediated by biochemically distinct SWI/SNF complexes. *PLoS Genetics*, *11*, e1005748. http://dx.doi.org/10.1371/journal.pgen.1005748.

Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy in Science*, *61*, 297–312.

Salmon, W. C. (1997). Causality and explanation: A reply to two critiques. *Philosophy in Science*, *64*, 461–477.

Sandell, L. L., & Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery, and chromosome loss. *Cell*, *75*(4), 729–739.

Sarkar, S. (1998). *Genetics and reductionism*. Cambridge, UK; New York: Cambridge University Press.

Sathasivan, K., Haughn, G. W., & Murai, N. (1991). Molecular basis of imidazolinone herbicide resistance in Arabidopsis thaliana var Columbia 1. *Plant Physiology*, *97*, 1044–1050.

Schumer, M., Cui, R., Rosenthal, G. G., & Andolfatto, P. (2015). Reproductive isolation of hybrid populations driven by genetic incompatibilities. *PLoS Genetics*, *11*(3), e1005041.

Sober, E. (1980). Evolution, population thinking, and essentialism. *Philosophy of Science*, *47*, 350–383.

Spencer, F., & Hieter, P. (1992). Centromere DNA mutations induce a mitotic delay in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *89*, 8908–8912.

Sterelny, K., & Kitcher, P. (1988). The return of the gene. *The Journal of Philosophy*, *85*(7), 339–361.

Stern, D. L. (2000). Evolutionary developmental biology and the problem of variation. *Evolution: International Journal of Organic Evolution*, *54*, 1079–1091.

Stern, D. L. (2010). *Evolution, development, and the predictable genome*. Greenwood Village, CO: Roberts & Company Publishers.

Stern, D., & Orgogozo, V. (2008). The loci of evolution: How predictable is genetic evolution? *Evolution: International Journal of Organic Evolution*, *62*, 2155–2177. http://dx.doi.org/10.1111/j.1558-5646.2008.00450.x.

Stoye, J. P., & Yap, M. W. (2008). Chance favors a prepared genome. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 3177–3178. http://dx.doi.org/10.1073/pnas.0800667105.

Tassabehji, M. (2003). Williams–Beuren syndrome: A challenge for genotype–phenotype correlations. *Human Molecular Genetics*, *12*(Suppl. 2), R229–R237.

Versluis, D., D'Andrea, M. M., Garcia, J. R., Leimena, M. M., Hugenholtz, F., Zhang, J., et al. (2015). Mining microbial metatranscriptomes for expression of antibiotic resistance genes under natural conditions. *Scientific Reports*, *5*, 11981.

Wade, N. (2003). Gene sweepstakes ends, but winner may well be wrong. *New York Times* www.nytimes.com/2003/06/03/science/03GENE.htm.

Ward, L. (1923). The genetics of curly wing in Drosophila. Another case of balanced lethal factors. *Genetics*, *8*, 276.

Waters, C. K. (2007). Causes that make a difference. *Journal of Philosophy*, *104*, 551–579.

Weber, M. (2005). *Philosophy of experimental biology*. New York, NY: Cambridge University Press.

Weischenfeldt, J., Symmons, O., Spitz, F., & Korbel, J. O. (2013). Phenotypic impact of genomic structural variation: Insights from and for human disease. *Nature Reviews Genetics*, *14*, 125–138.

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. New York; Oxford: Oxford University Press.

Zachar, Z., & Bingham, P. M. (1982). Regulation of white locus expression: The structure of mutant alleles at the white locus of Drosophila melanogaster. *Cell*, *30*(2), 529–541.