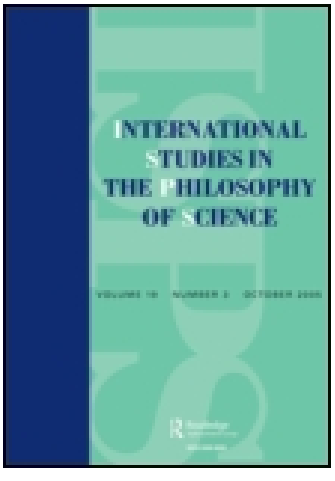


International Studies in the Philosophy of Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cisp20

The Referential Convergence of Gene Concepts Based on Classical and Molecular Analyses

Tudor M. Baetu
Published online: 18 Feb 2011.

To cite this article: Tudor M. Baetu (2010) The Referential Convergence of Gene Concepts Based on Classical and Molecular Analyses, International Studies in the Philosophy of Science, 24:4, 411-427

To link to this article: http://dx.doi.org/10.1080/02698595.2010.543351

International Studies in the Philosophy of Science
Vol. 24, No. 4, December 2010, pp. 411–427

Routledge
Taylor & Francis Group

The Referential Convergence of Gene Concepts Based on Classical and Molecular Analyses

Tudor M. Baetu

Kenneth Waters and Marcel Weber argue that the joint use of distinct gene concepts and the transfer of knowledge between classical and molecular analyses in contemporary scientific practice is possible because classical and molecular concepts of the gene refer to overlapping chromosomal segments and the DNA sequences associated with these segments. However, while pointing in the direction of coreference, both authors also agree that there is a considerable divergence between the actual sequences that count as genes in classical genetics and molecular biology. The thesis advanced in this paper is that the referents of classical and molecular gene concepts are coextensive to a higher degree than admitted by Waters and Weber, and therefore coreference can provide a satisfactory account of the high level of integration between classical genetics and molecular biology. In particular, I argue that the functional units/cistrons identified by classical techniques overlap with functional elements entering the composition of molecular transcription units, and that the precision of this overlap can be improved by conducting further experimentation.

1. Introduction

Classical geneticists defined genes as 'difference makers' in respect to phenotypes—for example, as alleles associated with alternative phenotypes found at a given chromosomal locus (Morgan 1935; Moss 2003; Rheinberger and Müller-Wille 2008; Waters 1994). Thomas Hunt Morgan and his collaborators (Morgan et al. 1915; Sturtevant 1913) mapped genes on linear genetic linkage maps, and, in conjunction with the Sutton-Boveri chromosomal interpretation, cytological studies showed that genetic maps correspond with the order of the genes along chromosomes (Bridges 1916; McClintock 1929; Painter 1934). It is now generally acknowledged that classical genetics

Tudor Baetu is at the Department of Philosophy, University of Maryland. Correspondence to: Department of Philosophy, University of Maryland, 1125D Skinner Building, College Park, MD 20742, USA. E-mail: tbaetu@hotmail.com

ISSN 0269–8595 (print)/ISSN 1469–9281 (online)
DOI: 10.1080/02698595.2010.543351

succeeded in explaining some characteristics of inheritance (e.g. segregation, linkage, recombination) by hypothesizing the existence of genes arranged in a linear fashion along chromosomes (Darden 1991).

One of the immediate reactions to the elucidation of the structure of DNA was the identification of genes with DNA sequences (Watson and Crick 1953). Shortly after, the Central Dogma of molecular biology allotted DNA the role of an archive containing information eventually duplicated and used to somehow determine phenotypes via yet to be elucidated mechanisms (Crick 1958). During the 1960s and 1970s, the notion that genes are DNA sequences coding information for protein synthesis (the gene as a 'blueprint' for protein synthesis) was reframed in terms of syntax-like conserved motifs providing instructions for transcription and translation (reviewed in Gerstein et al. 2007). The elucidation of the mechanisms of gene expression strongly suggested that what makes the difference between a DNA sequence that eventually contributes to a phenotype and one that does not hinges on the presence of conserved sequences responsible for recruiting the transcriptional and translational machinery of the cell. Most notably, modern molecular concepts define genes as regulatory regions preceding coding sequences (the transcription unit concept), or as fragments of translated DNA/collinear with a peptide sequence (the open reading frame concept).

While we may be tempted to view classical genetics and molecular biology as historical currents that succeeded one another, this view is not adequate. Lindley Darden (2006, 98) convincingly argues that classical genetics and molecular biology elucidate 'separate but serially connected mechanisms'. According to this account, Gregor Mendel provided a highly schematic outline of a series of events explaining inheritance phenomena. Then, classical genetics elucidated in more detail some elements contained in this scheme (e.g., the mechanism of allelic segregation, explained by meiosis, and recombination, explained by chromosomal crossing over) while relegating other elements to 'black boxes' (the ability of alleles to replicate and determine phenotypes). Finally, molecular biology gradually filled in the remaining 'black boxes' with more and more mechanistic details (most famously, the mechanisms of DNA replication and gene expression), until the present-day picture of genetics emerged. The 'serially connected mechanisms' account specifies the aspects of inheritance explained by classical genetics and those explained by molecular biology, and shows how the mechanisms postulated by classical and molecular explanations fit together without leaving gaps in the productive continuity from start (the genotype of the parents) to finish conditions (the expression of specific traits in the offspring).

It is also important to realize that contemporary scientific practice integrates discovery strategies, experimental techniques and concepts inherited from both classical genetics and molecular biology. Contemporary research practice in genetics is driven by two main discovery strategies. Classical-style analysis (forward genetics, genetic studies) starts at the level of an inherited phenotype, then genes responsible for the phenotype are hypothesized and attempts are made to localize the genes as chromosomal loci, and eventually as DNA sequences. Molecular analysis (cloning, reverse genetics) follows the reverse pattern. Genomic DNA is sequenced, putative genes are identified by searching for potential transcription units and open reading

frames using computer algorithms such as BLAST. Preliminary research then shows that the putative genes can be expressed in artificial systems such as plasmid transfected cells, and, eventually, the DNA sequences in question are shown to contribute to a phenotype or biological function via the generation of transgenic/knockout organisms. Both discovery strategies and their associated experimental techniques are often intimately intertwined in the same study. Several philosophers of biology (Falk 2003; Vance 1996; Waters 2008; Weber 2005) point out that contemporary research in genetics combines classical experimental and explanatory strategies, such as breeding and genetic linkage mapping of phenotypes, with cloning, sequencing and reverse genetic analysis introduced by molecular biology. Furthermore, documented cases of 'explanatory interference'—cases when the accommodation of data from molecular biology results in a more precise genotyping and more adequate classical-style explanations of the transmission patterns associated with an inherited condition—hints to a high level of integration of the two sciences as a unique field of research (Baetu 2011).

Finally, it is worth noting that the Human Genome Nomenclature Committee (HGNC) defines a 'gene' as 'DNA segment that contributes to phenotype/function. In the absence of a demonstrated function a gene may be characterized by sequence, transcription or homology' (Wain et al. 2002, 464). This definition encompasses both classical and molecular gene concepts. The first half of the HGNC definition accommodates a situation where a segment of DNA is shown to be associated with a phenotype even if this segment is too short (e.g., a point mutation) or too long (e.g., a large DNA segment found between two chromosomal break points) to count as a molecular gene. At the same time, the second half of the HGNC definition is explicitly meant to accommodate a situation where putative genes are identified via genome annotation techniques (e.g., genome annotation algorithms, homology with gene product sequences and known genes).

Treating classical genetics and molecular biology as scientific currents that succeeded one another raised traditional philosophical questions about the growth and continuity of scientific knowledge from one theory to the next. Answering such questions is at the origin of the well-known reductionism vs. antireductionism debate in genetics (Kitcher 1984; Waters 1990). Realizing the current status of integration between classical genetics and molecular biology prompts however a different, more pressing question: How is it possible to use side by side gene concepts issued from classical genetics and molecular biology? Furthermore, given the tight relationship between gene concepts and the discovery strategies and experimental techniques used to identify them, how is it possible to transfer knowledge about genes acquired by means of classical-style analysis in the context of molecular biology?

These questions received insufficient attention in the philosophical literature. Kenneth Waters and Marcel Weber argue that the joint use of distinct gene concepts and the transfer of knowledge between classical and molecular analyses is possible because classical and molecular concepts of the gene refer to overlapping chromosomal segments and the DNA sequences associated with these segments. Waters (1994, 2004) proposes an unification account according to which both concepts refer to differences

in DNA sequences, while Weber (2005) argues that, at least in some cases, classical and molecular concepts pick overlapping DNA segments. Weber further exemplifies how such an extensional overlap explains the transfer of knowledge about genes from classical genetics to molecular biology.

Unfortunately, things are not so simple. While pointing in the direction of coreference, both authors also agree that there is a considerable divergence between the actual sequences that count as genes in classical genetics and molecular biology. For Waters, the overlap can be as minimal as a point mutation within a transcription unit spanning thousands of base pairs, while Weber argues that sometimes there might be no overlap at all, taking this as an indication that classical and molecular concepts refer to distinct natural kinds. In the end, the reader cannot help noticing that most of what Waters and Weber give with one hand in guise of a solution, they take back with the other. It is not at all clear that the little that is left suffices to answer any of the above raised questions. If anything, the results of their analyses add an even more puzzling question: If the amount of overlap between the sequences picked by classical and molecular concepts is not only minimal, but also absent in some situations, what justifies the confidence with which contemporary researchers rely on both concepts?

The thesis advanced in this paper is that the referents of classical and molecular gene concepts are coextensive to a higher degree than admitted by Waters and Weber, and therefore coreference can provide a satisfactory account of the high level of integration between classical genetics and molecular biology.

The paper is organized as follows. In Section 2, I analyse Waters's and Weber's accounts of the relationship between classical and molecular gene concepts in respect to the issue of referential continuity. In Section 3, I provide arguments showing that not all classical gene concepts must be taken into consideration; it suffices that we retain the functional unit/cistron concept. This allows me to conclude that the overlap between classical and molecular genes spans extended segments of chromosomal DNA. Second, I show that the resolution of functional units, as defined by classical analysis, is dependent on the number of distinct complementation experiments. This allows me to argue that the precision of the overlap between classical and molecular genes depends on how far complementation experiments have been pushed. Third, I show that classical 'difference makers' are located somewhere within a transcription unit, and, more importantly, classical functional units correspond to functional elements within molecular transcription units. This shows that there is an overlap between classical and molecular genes. In the concluding section, I summarize my findings and arguments.

2. Accounts of the Coreference of Classical and Molecular Gene Concepts

2.1. The 'Unifying Molecular Gene Concept' Account

Philosophers of biology were quick to note that, from an experimental point of view, classical genetics aims to tie differences in phenotypes to differences in genotype. What mattered primarily to classical geneticists was the relationship between a change in a

gene and a change in phenotype, rather than the nature of these entities themselves; this is known as the 'difference maker' concept of the gene (Moss 2003, 46; Rheinberger and Müller-Wille 2008, 6; Waters 1994, 169). Waters notes that

> 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. (Waters 1994, 182–183)

The conceptions contrasted here are the classical 'difference maker' concept and the molecular open reading frame (sequences translated into protein gene products) concept. Since the former can refer to single base mutations in both coding and regulatory sequences, while the latter usually refers to extended DNA sequences counting as coding regions or open reading frames, Waters concludes that the referents of classical and molecular gene concepts can be significantly different.[1]

Nevertheless, Waters argues, the molecular gene 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. … The differences by which classical genes were identified are taken to be differences in nucleotide sequences affecting the transcription of molecular genes. (Waters 1994, 163, 183)

The argument at work here is that molecular biology establishes that differences in genotypes associated with differences in phenotype (i.e., classical genes) are due to differences in DNA sequences. The molecular analysis of the material found at the chromosomal loci associated with mutant and wild-type alleles, as defined by classical analysis, reveals a difference in DNA sequence. Furthermore, the experimental manipulation of the genetic material shows that the phenotype changes from wild-type to mutant and vice versa by making the appropriate changes in DNA sequence. Thus, talk about classical 'difference makers' can be replaced by talk about mutations, deletions and insertions in DNA sequences associated with a given chromosomal locus.[2]

2.2. The 'Floating Reference' Account

Weber shows how classical gene concepts were used to identify molecular genes. He explains, for instance, how molecular biologists knew they identified the same *white* gene Morgan investigated previously by means of classical techniques (2005, 216–217).[3] On the basis of such examples, Weber concludes that, at least in some cases, the classical and molecular analysis based gene concepts pick overlapping DNA segments: 'gene concepts are united by an extensional overlap between the different stages of the term's historical development' (Weber 2005, 204). At the same time, in a manner reminiscent of Waters's conclusion, Weber also argues that, although coextensive in many cases, the referents of the two concepts may fail to converge. Based on the example of the cloning of the *achaete–scute* locus, Weber (2005, 220) concludes that 'in this case the classical genetic and the molecular analysis did not correspond'. In order to accommodate the conflicting evidence, showing that sometimes the classical and molecular

concepts corefer while at other times they fail to do so, Weber proposes a 'freely floating reference' account:

> As the practice of genetics continuously generated new ways of detecting, localizing (mapping), and describing genes, some DNA segments moved in, others out of the term's extension. (Weber 2005, 224)

Unfortunately, Weber does not discuss in more detail the experimental constraints that determine the degree of precision to which DNA sequences picked by classical and molecular gene concepts overlap. Nor does he discuss the implication that classical and molecular concepts fail to corefer in a consistently reliable way. Instead, he chooses to pursue a different avenue, by drawing a connection between the variable coextension of the classical and molecular concepts and the absence of a unique natural kind called 'gene':

> the natural kinds denoted, for example, by the classical gene concept of the Morgan school or the neoclassical concept of microbial geneticists, are different from the natural kinds denoted by more recent, molecularized versions of the gene concept. … on my account of the gene concept, different historical versions of this concept did refer, but to different natural kinds that are not coextensive. My analysis also reveals that the extensions of different versions of the gene concept were not completely disjoint. … several subtypes of the natural kind picked out by the classical gene concept also lie in the extension of the molecular gene concept. (Weber 2005, 225–226)

3. A Revised Account of the Referential Convergence of Classical and Molecular Concepts

I argue that neither Waters's nor Weber's account does full justice to the current status of integration between classical genetics and molecular biology. The first problem is the failure to take into consideration the fact that the functional unit (cistron) concept replaced previous concepts within classical genetics prior to the birth of molecular biology. This is an extremely important point, since the adoption of functional units explicitly entails a much more significant overlap that extends beyond point mutations located somewhere within transcriptional units. Also, the proposed accounts fail to take into consideration the experimental constraints that determine the degree of precision of this overlap. I will show that the identification of functional units is subjected to experimental constraints that can explain the variable degree of overlap between the sequences picked by classical and molecular concepts. Thus, there is no need to attribute this variability to a variety of distinct kind referents.

A second problem is that both authors dismiss the transcription unit concept in favour of the open reading frame concept. Weber discusses in detail the use of complementation in the discovery of open reading frames; however, he doesn't say anything about complementation as a tool in the investigation of the transcriptional regulation of genes (which is a standard textbook topic: Griffiths et al. 2005). As I will show in this section, not only are classical 'difference makers' always located somewhere within a transcription unit, but functional units correspond to functional

elements within transcription units. This is a significant result since it shows that we don't have to worry about cases of divergent reference and complete lack of overlap.

3.1. Both Classical and Molecular Gene Concepts Refer to Extended Segments of DNA

Waters's account supports referential continuity in one crucial respect: both classical and molecular gene concepts ultimately point to DNA sequences. On the other hand, however, classical 'difference makers' refer indiscriminately to point mutations, deletions, insertions, and chromosomal crossing-overs, while from a molecular perspective not any collection of nucleotides can be transcribed and translated into gene products, and therefore contribute to a phenotype. A similar critique applies to Weber. The 'floating reference' account gives equal credit to a variety of concepts, even if, collectively, they only loosely refer to overlapping DNA sequences.

I argue that classical and molecular gene concepts are coextensive not only in the general sense that classical 'difference makers' point to differences in DNA sequence that can be investigated by molecular techniques, but also in the more specific sense that classical and molecular concepts refer to overlapping segments of DNA.

Discovered as early as 1911 (Morgan 1911), complementation refers to a situation whereby the crossing of two strains homozygous for different mutations associated with the same mutant phenotype yields a wild-type phenotype (Benzer 1955; Lewis 1951). Classical geneticists interpreted complementation as an indication that the two mutations affect two distinct 'genetic units', dubbed 'functional units' (or cistrons). If the mutations were in the same unit, then the offspring could have not received a copy of the wild-type unit, since none of the parents had one to begin with. Low-frequency occurences of complementation were attributed to intragenic recombination events, and were used in order to generate high-resolution genetic maps. Complementation and recombination data enabled classical geneticists to map with various degrees of precision phenotypes as a collection of functional units spanning a certain length along chromosomes (Figure 1).

Complementation provided evidence that (i) genes are not, or are not associated with, point locations on chromosomes (such as units of mutation or recombination), but rather extended chromosomal segments. Complementation also showed that (ii) more than one functional unit is typically required for the expression of any given phenotype. Furthermore, since mutations targeting apparently unrelated traits can complement, it also turned out that, in many cases, (iii) a single mutation in one gene can affect several traits/biological functions. Finally, since non-complementing mutations can map at distinct chromosomal loci (iv) it is possible to distinguish between mutations in the same functional unit that result in an identical phenotype (in other words, classical geneticists were able to distinguish mutations that cannot be differentiated by observing phenotypes before DNA sequencing became available).

Finding (i) indicates that the 'difference maker' concept does not provide a sufficiently specific characterization of the classical gene. Prior to the discovery of complementation, classical geneticists may have very well used other types of genetic 'difference makers' for a variety of experimental and explanatory purposes—most

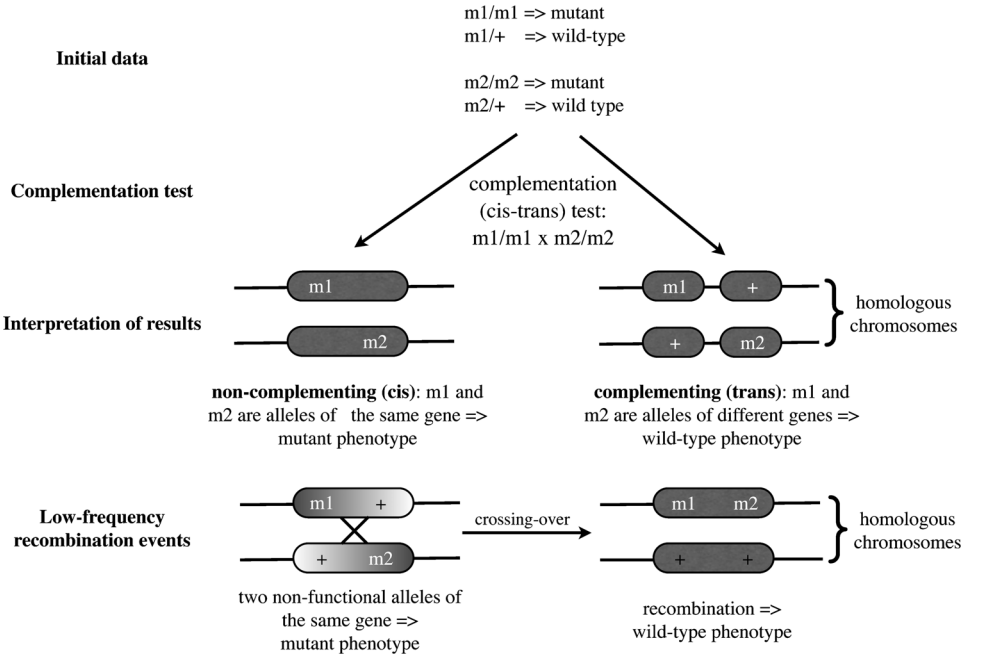


Figure 1 Complementation.

notably mutation units and recombination units (determining the resolution of genetic linkage maps), both of which amount to point locations corresponding, depending on the resolution of the techniques used, to one or two DNA base-pairs. However, with the establishment of the Morgan school, geneticists became aware that 'difference makers' are organized in functional units spanning well-defined lengths on genetic and chromosomal maps. Findings (ii) and (iii) indicated that the simplistic one gene–one phenotype must be abandoned. Finding (iv) indicates that classical geneticists were able to discover 'difference makers' that didn't make a difference in terms of phenotype.[4]

I conclude that we don't have to consider all the gene concepts inspired by classical analysis; it suffices that we retain the functional unit concept. The view that genes correspond to point locations rather than extended segments was invalidated within classical genetics because of the discovery of complementation.[5]

3.2. Classical and Molecular Concepts Can Refer to Exactly the Same Sequences

The next task is to determine to what extent functional units overlap with molecular genes. It is important to realize that the adoption of the functional unit concept brought about a great number of experimental challenges. As new mutants were created in the lab, classical geneticists soon discovered that in more and more cases the same difference in phenotype 'maps' at distinct chromosomal loci. Yet whether distinct functional units are associated with these loci is a matter of further investigation. If the

two loci are close to each other, they could count as two variants (alleles) of the same unit, or they may involve distinct units located next to each other; if they are far apart, they may be two redundant copies of the same unit, or they may determine their associated phenotype in a very different way. Knowledge about the chromosomal loci associated with a given phenotype does not allow for unambiguous inferences about the number of functional units involved (establish whether two loci span the same gene) and for an unambiguous identification of functional units (establish whether the same or a different unit is found at more than one chromosomal locus). Thus, genetic and chromosomal mapping of phenotypic mutations turned out to be insufficient; conducting complementation assays on a regular basis was also necessary.

Any single complementation test can only decide if two mutations affect the same or different functional units. Complementation is thus sufficient to map genes as extended segments of genomic material, but not enough to establish their precise length and location onto chromosomes and chromosomal DNA. It takes a great number of recombination events between a large number of mutations in order to map a functional unit as an extended segment of genomic DNA of a definite length (i.e., a cistron). This degree of precision was eventually achieved by Seymour Benzer (1955). It is now generally acknowledged that Benzer's map of the *rII* region of the T4 phage (a virus infecting bacteria) helped to unite molecular biology and classical genetics because of its unprecedented high resolution:

> [It is possible to estimate] the frequency of recombination between markers separated by a distance of one nucleotide pair. This is not less than 5×10^{-6} since the maximum size of he *rII* gene is 1.6×10^{4} nucleotide pairs, which corresponds to 8 percent recombination. Since the present procedure allows the detection of wild type recombinants in proportions as low as 10^{-8}, it is feasible to establish this fundamental frequency experimentally. (Benzer, quoted in Holmes 2006, 236)

The tremendous significance of Benzer's map is that while units of mutation and recombination show variable degrees of overlap with transcription units and open reading frames—ranging from one nucleotide within an open reading frame to several open reading frames contained in a chromosomal fragment subjected to insertion, deletion or crossing over—cistrons were eventually shown to provide a quasi-perfect overlap with open reading frames subsequently identified by molecular techniques (Mosig and Eiserling 2006). This is a very encouraging result, since it shows that, at least in some cases, classical and molecular concepts refer to the same sequences.

3.3. Experimental Constraints Determining the Precision with Which the Sequences Picked by Classical and Molecular Concepts Overlap

Encouraging as the above conclusion may be, the fact remains that the degree of overlap between sequences identified via classical and molecular techniques is usually less than perfect. There are two aspects to be considered here. First, there is a resolution problem: classical and molecular concepts refer to roughly the same chromosomal segments and their associated DNA sequences rather than precisely the same fragments/sequences. Second, the discovery that mutations in regulatory sequences

(sequences involved in transcription regulation never translated into protein gene products) can complement mutations in their associated coding sequences/open reading frames indicates that there is no necessary correspondence between functional units and open reading frames. I will begin by addressing the first problem.

Benzer's ability to produce genetic maps at a resolution equivalent to that of single nucleotides depended largely on the peculiarities of the system he used: phages infecting bacteria. Phages can be produced and managed in large numbers by growing them in an equally manageable host selection system. Equally important, phages have a very short genome. Given a sufficiently high number of mutations (Benzer identified approximately 2,400 mutations distributed between two functional units, *rII*A and *rIIB*), it becomes more and more probable that the mutations target all the nucleotides of the genome. Given the possibility of detecting the lowest frequency of recombination between these mutations (Benzer observed a lowest frequency of 0.02%, though the system had a resolving power of 0.0001%), it becomes more and more probable that recombination occurs between adjacent nucleotides. A high number of mutations, paired with the detection of lowest frequency of recombination, ensures that the genetic linkage map coincides with the DNA sequence of the viral genome.

In contrast, if 10^9 phages is a manageable and sufficient number when it comes to mapping the 170,000 base-pairs of the T4 phage, 10^9 flies is both unmanageable and insufficient when it comes to mapping the 165 million base-pairs of the *Drosophila* genome. Experimental constraints dictate that the resolution of genetic maps generated by classical techniques is bound to be lower in bacterial genetics, considerably lower in *Drosophila* genetics, even lower in murine genetics and extremely low in human genetics. Set aside this general rule of thumb, we also have to keep in mind that the frequency of recombination varies along the chromosome; this is especially true of the linear eukaryotic chromosomes (recombination events being more rare closer to the centromeres, or centre, and telomeres, or ends, of the chromosomes).

Given this dependence of the resolution of genetic maps on experimental constraints, it can be argued that the variable overlap between the sequences picked by classical and molecular concepts in higher organisms does not prove that we are dealing with distinct natural kinds.[6] It could be due to the lack of data points.[7] In fact, in one of the review articles cited by Weber as a scientific source for his discussion of the *achaete–scute* locus, Alain Ghysen and Christine Dambly-Chaudiere explicitly claim that

> One of the outcomes of the molecular analysis of AS-C [*achaete–scute*] is that the phenotype of most *scute* alleles can now be correlated with their location on the chromosome … This could not be done before, because the incidence of recombination in this region is too low to permit the meiotic mapping of alleles within AS-C. (Ghysen and Dambly-Chaudiere 1988, 495–496)

Maria Alonso and Carlos Cabrera, the researchers that analyzed the DNA sequence of the *achaete–scute* locus (also cited by Weber), further point out that 'As the AS-C is located at the tip of the X chromosome … classical genetic analysis is hampered by the low recombination frequencies' (Alonso and Cabrera 1988, 2588).

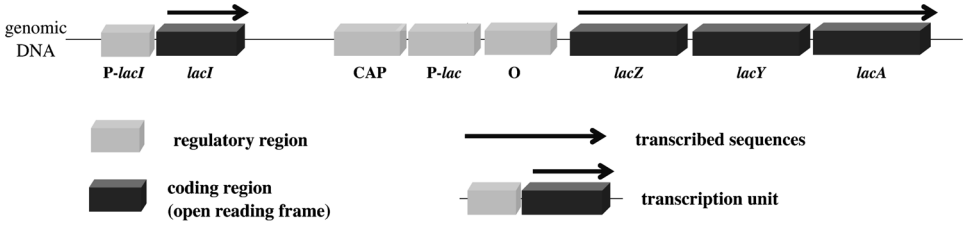
3.4. Classical Functional Units Pick Sequences Corresponding to Functional Elements Composing Molecular Transcription Units

In regard to the second issue, concerning the lack of correspondence between functional units and open reading frames, I want to point out that there is a consistent overlap of functional units with transcription units.

As it turns out, it was eventually established that complementation assays may lead to the identification of chromosomal segments that do not overlap with open reading frames no matter how manageable the organism model is and how extensively experimentation is conducted; this is, in part, what Weber aims to show in his discussion of the *achaete–scute* example. Classical genetics is blind to the mechanisms linking genotype and phenotype, weakness that renders it unapt to account for certain anomalies. Retrospectively, we know now that complementation is possible when recessive loss of function mutations occur in genes responsible for different steps in the same metabolic or signal transducing pathway. Complementation tests cannot handle 'dominant negative' mutations (loss of function mutations that are dominant rather than recessive). Furthermore, effects due to chromosomal pairing, such as transvection (when the enhancer on one chromosome affects transcription of the copy of the homologous gene situated on the other chromosome, Duncan 2002) and intragenic complementation (when a protein has multiple independently mutable domains, Ullmann, Jacob, and Monod 1967, or is composed of several subunits, Garen and Garen 1963) hinder a researcher's ability to design, conduct, and interpret complementation experiments.

The possibility that the DNA segments identified via classical and molecular analysis may not perfectly coincide is a real concern; as discussed earlier, the degree of overlap is highly dependent on experimental constraints. However, the possibility that they may not overlap at all is not. In order to prove my point, I will begin by making explicit a corollary of Waters's analysis. If differences in genotype responsible for phenotypic differences established via classical analysis are in fact differences in DNA sequences that affect gene expression, it must be the case that classical 'difference makers' are sequence differences within transcription units. Differences in DNA sequence outside coding and regulatory sequences, that is, DNA sequences that are never transcribed/translated and are not recognized by DNA/RNA binding factors playing a role in gene expression regulation are not associated with differences in phenotype and, therefore, cannot count as classical 'difference makers'. Thus, differences in DNA sequences responsible for phenotypic differences (i.e., classical genes) are always located within the boundaries of transcription units (i.e., molecular genes). This conclusion is supported by the contemporary practice documented by both Weber (2005) and Waters (2008): classical analysis indicates that the DNA associated with a certain locus contributes to a phenotype, while molecular analysis delimits the DNA sequences (regulatory and coding) that need to be present in order for DNA to be expressed and contribute to a certain phenotype.

Similarly, in regard to cases where complementation assays fail to delimit functional units that coincide with open reading frames no matter how extensively these assays are performed, the concern is not that classical analysis may identify random DNA segments



Sequence Element	Function
P-*lacI*	promoter of *lacI* (binding site for RNA polymerase)
lacI	encodes LacI (transcriptional repressor protein)
CAP	binding site for Catabolite Activator Protein (protein acting as a transcriptional activator)
P-*lac*	promoter of the *lac* operon (binding site for RNA polymerase)
O	operator (binding of LacI at O blocks the binding of RNA polymerase at P-*lac*)
lacZ	encodes β-galactosidase (enzyme that cleaves lactose into glucose and galactose)
lacY	encodes β-galactoside permease (protein transporting lactose across the cell membrane)
lacA	encodes β-galactoside transacetylase (role in lactose metabolism not yet understood)

Figure 2 Functional elements of the *lac* operon identified by means of complementation assays.

(e.g., stretches of 'junk' DNA that don't play any regulatory or coding role). Rather, the concern is that classical analysis may identify a fragment of a transcriptional unit (e.g., a functional coding domain or regulatory region) instead of an open reading frame. This constitutes a shortcoming only in as much as complementation is seen as a strategy for discovering open reading frames. In contemporary practice, complementation is used primarily to investigate the regulation of transcription. One of the most famous examples is the elucidation of the regulatory mechanisms of the *lac* operon (Jacob and Monod 1961). In this example, the functional units identified by classical techniques did not always coincide with open reading frames, but they did consistently pick functional motifs composing molecular transcription units, most notably minimal promoter sequences required for transcription, binding sites for regulatory proteins, and open reading frames located in the coding sequence of transcription units (Figure 2).

Even in Weber's alleged example of referential divergence, the classical analysis of the *achaete–scute* locus revealed the presence of three open reading frames (*achaete*, *scute*, *lethal of scute*). As for the remaining complementation groups (*scute* β and γ), they are now thought to correspond to distal, alternative regulatory elements affecting the transcription of *scute* (Ghysen and Dambly-Chaudiere 1988, 499). Thus, all of the classical functional units pointed in the end to functional molecular elements involved in the regulation and expression of the locus, that is, to functional elements composing molecular transcription units.[8]

4. Conclusion

I conclude that the DNA sequences picked by classical and molecular concepts overlap to a higher extent than acknowledged by current accounts. Contra Waters, I argued that

classical 'difference makers' are located within transcription units. Furthermore, the view that genes correspond to point locations rather than extended chromosomal segments was invalidated within classical genetics because of an incompatibility with the results of complementation assays. Classical analysis aims to identify genetic 'difference makers' organized in extended segments of DNA that behave like functional units (rather than 'difference makers' *simpliciter*). Thus, classical and molecular gene concepts must minimally refer to overlapping DNA sequences spanning more than one base-pair.

Contra Weber, I argued that the DNA sequences picked by classical and molecular concepts overlap and that the variable overlap between these sequences does not stem from fundamental reference continuity incompatibilities, but from experimental limitations. I argued that the precision and resolution with which functional units are mapped onto chromosomal DNA is proportional to the number of individual complementation data points. I also showed that, although it is true that classical functional units don't always correspond to open reading frames, they do correspond to functional elements within molecular transcription units (e.g., regulatory sequences, open reading frames, functional domains). Finally, I used these findings in order to defuse his example of referential divergence.

The picture emerging from my analysis is that of two concepts referring to overlapping DNA sequences. Classical concepts in general pick DNA sequences falling within the boundaries of molecular transcription units, while classical functional units specifically refer to functional DNA-sequence elements entering the composition of molecular transcription units. Such a convergence does not entail a tight conceptual unity, yet provides sufficient means to integrate the findings of classical and molecular genetics as part of the same body of information about inheritance phenomena (e.g., devise 'interfield' explanations, share techniques, guide analysis). Referential continuity enables molecular biologists to continue to make use of classical techniques and to use the results already obtained by means of these techniques in conjunction with the newly developed molecular theoretical framework and associated experimental techniques. In this respect, it is worth noting that it makes no sense to mix classical and molecular gene concepts and their associated discovery strategies, if genes, as defined by classical and molecular concepts, refer to different entities. If scientists rely on both concepts, this is because they assume that both point to overlapping DNA sequences. Given past successes (e.g., the mapping of the *rIIA* and *rIIB* genes, the *lac* operon, the cloning of the *white* gene, and other textbook examples), this assumption is a partially justified working hypothesis. Furthermore, this assumption has proved so far to be a successful strategy (e.g., the elucidation of the *tuf-s10* operon in *Borrelia burgdorferi*, which causes Lyme disease; the cloning of the human growth hormone receptor; complementation of *env* region in HIV).

Genetics provides an interesting example of referential continuity that may find parallels in other cases. Typically, concepts refer or fail to refer to the same entities. However, in the case of genetics, the referent is a stretch of DNA located at a given chromosomal locus. Thus, it is not only question of establishing that classical and molecular concepts refer to DNA and places on chromosomes, but it must also be established the degree of precision with which these DNA sequences and chromosomal loci overlap. It

is interesting to note that while the sequences/loci picked by classical concepts changed over time, the change followed the direction of an increase in the specificity and precision of the overlap, eventually matching the resolution offered by molecular concepts. The few historical cases in which the sequences picked by classical concepts reached the same level of resolution as molecular terms, and were eventually shown to coincide with molecular genes, have played a crucial role in strengthening the scientists' conviction that classical and molecular gene concepts corefer. From there on, this conviction, which is nothing else than a partially justified working hypothesis, has proved to be a very fruitful assumption that gained has further confirmation with each additional success.

Acknowledgements

I would like to thank Lindley Darden, Erika Milam, Eric Saidel, Pamela Henson, Joan Straumanis, Elizabeth Schechter, Brendan Ritchie, the DC History and Philosophy of Biology group, as well as two anonymous referees of this journal for helpful discussion and comments on earlier drafts. This work was supported by the Fonds de la recherche sur la société et la culture, Québec, Canada, grant number 127231.

Notes

[1] Waters initially argued that, in order to count as a gene, a segment of DNA must be 'biochemically active' (Waters 1990, 130). According to this conception, sequences required for recruiting the transcriptional machinery are part of the gene, suggesting that the DNA sequences to which Waters refers include open reading frames, as well as transcription units. In a subsequent development, Waters argues that molecular genes are sequences of DNA coding for product sequences generated at some point during gene expression (Waters 1994, 178). One peculiarity of this homology-based conception is that only coding DNA counts as genes, while regulatory regions (promoters, enhancers) are left out the gene; this entails that the concept refers to open reading frames, but not to transcription units.

[2] Differences in DNA sequence can be investigated by molecular techniques, hence supporting Waters's (2008) subsequent view that molecular biology contributed to an experimental/methodological 'retooling' of classical genetics.

[3] The position of the gene was narrowed down by classical analysis, via cytological band mapping on polytene salivary gland chromosomes in *Drosophila* (giant, easily observable chromosomes formed by multiple DNA replications without mitosis). Then, transcription units and open reading frames (i.e., molecular genes) were identified by sequencing and annotating DNA fragments binding specifically the same position on the chromosome. By inserting the mutant or the wild-type version of the molecular gene, researchers were able to produce the same differences in phenotype associated with the mutant and wild-type allele of the classical gene, thus proving that the classical and molecular genes referred to the same DNA sequences, located at the same chromosomal locus.

[4] In classical terms, such mutations are distinguished solely by different recombination frequencies (e.g., for the most part, Benzer worked only with the rapid mutant, phenotype caused by several distinct mutations in the rII region).

[5] The chromosomal fragments/DNA sequences that count as units of mutation and recombination vary with the resolution of the experimental techniques used. In this sense, Weber's (2005, 224) claim that the evolution of gene concepts was driven by the development of new

techniques is correct. However, I also want to point out that some concepts were irrevocably abandoned because of their incompatibility with new discoveries.

[6] I think that Weber's concern with natural kinds is blurring the issue. Appealing to the kinds of entities classical and molecular geneticists thought genes must be in order to fulfil certain explanatory roles makes it extremely difficult to provide an account of how it is possible for geneticists to use side-by-side concepts issued from classical genetics and molecular biology. For example, as Weber (2005, 209) himself argues, early geneticists thought genes have two distinctive properties: they are units, such that they can segregate and assort; and they are causes of phenotypic traits. Later, Muller (1951) conjectured that, in order to perform the role classical genetics ascribes them, genes must possess three properties: the abilities to replicate, to contribute to phenotypes and to mutate; he conjectured that genes are most likely proteins. After the elucidation of the structure of DNA, it became clear that DNA is a linear polymer unable to perform enzymatic activities; genes were then said to 'store information'. Finally, contemporary genetics defines genes in terms of structural motifs required for the recruitment of the gene expression machinery of the cell. If all these concepts refer, it follows that there are genes as causal entities, protein genes, blueprint genes and structural-motifs genes. Set aside the obvious problem that not all these entities really exist, it is not at all clear how such radically different kinds of entities could overlap extensionally. I think that the key to understanding the high level of integration between classical genetics and molecular biology lies in the fact that, leaving aside for a moment the various causal roles attributed to genes at various points in time, the classical and molecular concepts retained until today consistently include reference to a chromosomal locus and its associated DNA sequence. Thus, the real question is not about kinds, but about the extent to which the DNA sequences picked by classical and molecular gene concepts overlap.

[7] There are only six chromosomal break-points defining five functional regions of about 25,000 base-pairs each in the *achaete–scute* study (Weber 2005, fig. 7.2). This is a very small number compared to the thousands of recombination points analyzed by Benzer and is immediately reflected in the resolution of the map (25,000 vs. single nucleotides).

[8] Weber (2005, 227) acknowledges that 'if the regulatory regions are considered to be part of a molecular gene, then some molecular genes identified in *Drosophila* corresponded roughly to classical genes', but doesn't seriously consider this avenue of investigation. Yet regulatory sequences are part of the transcription unit concept (which can be seen as an open reading frame and its associated regulatory sequences). Not only is the latter an officially recognized and widely used concept (Wain et al. 2002 above; the NCBI GenBank guidelines), it also provides an indispensable theoretical framework for most genome annotation practices in use today: putative genes are identified on the basis of conserved promoter-like regulatory sequences preceding coding/transcribed/open reading frame sequences (Altschul et al. 1990; Mandoiu and Zelikovsky 2008).

References

Alonso, M., and C. Cabrera. 1988. The *achaete–scute* gene complex of *Drosophila* melanogaster comprises four homologous genes. *EMBO Journal* 7, no. 8: 2585–2591.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, no. 3: 403–410.

Baetu, T. M. 2011. Mechanism schemas and the relationship between biological theories. In *Causality in the sciences,* edited by P. McKay, J. Williamson, and F. Russo. Oxford: Oxford University Press.

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

Bridges, C. B. 1916. Non-disjunction as proof of the chromosome theory of heredity. *Genetics* 1: 1–53, 107–163.

Crick, F. H. 1958. On protein synthesis. *Symposia of the Society for Experimental Biology* 12: 138–163.
Darden, L. 1991. *Theory change in science: Strategies from Mendelian genetics.* New York: Oxford University Press.
——. 2006. *Reasoning in biological discoveries: Essays on mechanisms, interfield relations, and anomaly resolution.* Cambridge: Cambridge University Press.
Duncan, I. W. 2002. Transvection effects in *Drosophila*. *Annual Reviews of Genetics* 36: 521–556.
Falk, R. 2003. Linkage: From particulate to interactive genetics. *Journal of the History of Biology* 36, no. 1: 87–117.
Garen, A., and S. Garen. 1963. Complementation *in vivo* between structural mutants of alkaline phosphatase from *E. coli*. *Journal of Molecular Biology* 7: 13–22.
Gerstein, M. B., C. Bruce, J. S. Rozowsky, D. Zheng, J. Du, J. O. Korbel, O. Emanuelsson, Z. D. Zhang, S. Weissman, and M. Snyder. 2007. What is a gene, post-ENCODE? History and updated definition. *Genome Research* 17, no. 6: 669–681.
Ghysen, A., and C. Dambly-Chaudiere. 1988. From DNA to form: The *achaete–scute* complex. *Genes and Development* 2, no. 5: 495–501.
Griffiths, A., S. Wessler, R. Lewontin, W. Gelbart, D. Suzuki, and J. Miller. 2005. *Introduction to genetic analysis.* New York: W. H. Freeman.
Holmes, F. L. 2006. *Reconceiving the gene: Seymour Benzer's adventures in phage genetics.* New Haven, CT: Yale University Press.
Jacob, F., and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3: 318–356.
Kitcher, P. 1984. 1953 and all that: A tale of two sciences. *Philosophical Review* 93: 335–373.
Lewis, E. B. 1951. Pseudoallelism and gene evolution. *Cold Spring Harbor Symposia on Quantitative Biology* 16: 159–174.
Mandoiu, I., and A. Zelikovsky. 2008. *Bioinformatics algorithms: Techniques and applications.* Hoboken, NJ: Wiley.
McClintock, B. 1929. A cytological and genetical study of triploid maize. *Genetics* 14, no. 2: 180–222.
Morgan, T. H. 1911. The origin of five mutations in eye color in *Drosophila* and their modes of inheritance. *Science* 33: 534–537.
——. 1935. The relation of genetics to physiology and medicine. In *Les prix Nobel en 1933,* 1–16. Stockholm: Imprimerie Royale.
Morgan, T. H., A. H. Sturtevant, H. J. Muller, and C. B. Bridges. 1915. *The mechanism of Mendelian heredity.* New York: Henry Holt & Co.
Mosig, G., and F. Eiserling. 2006. T4 and related phages: Structure and development. In *The bacteriophages,* edited by R. Calendar and S. T. Abedon. Oxford: Oxford University Press.
Moss, L. 2003. *What genes can't do.* Cambridge, MA: MIT Press.
Muller, H. J. 1951. The development of the gene theory. In *Genetics in the 20th century: Essays on the progress of genetics during its first 50 years,* edited by L. C. Dunn, 77–99. New York: Macmillan.
Painter, T. S. 1934. A new method for the study of chromosome aberrations and the plotting of chromosome maps in *Drosophila melanogaster*. *Genetics* 19, no. 3: 175–188.
Rheinberger, H.-J., and S. Müller-Wille. 2008. Gene concepts. In *A companion to the philosophy of biology,* edited by S. Sarkar and A. Plutynski, 3–21. Malden, MA: Blackwell.
Sturtevant, H. 1913. The linear arrangement of six sex-linked factors in *Drosophila* as shown by their mode of association. *Journal of Experimental Zoology* 14: 43–59.
Ullmann, A., F. Jacob, and J. Monod. 1967. Characterisation by in vitro complementation of a peptide corresponding to an operator proximal segment of the beta-galactosidase structural gene of *Escherichia coli*. *Journal of Molecular Biology* 24: 339–343.
Vance, R. E. 1996. Heroic antireductionism and genetics: A tale of one science. *Philosophy of Science* 63 (Proceedings): S36–S45.
Wain, H. M., E. A. Bruford, R. C. Lovering, M. J. Lush, M. W. Wright, and S. Povey. 2002. Guidelines for human gene nomenclature. *Genomics* 79, no. 4: 464–470.

Waters, C. K. 1990. Why the anti-reductionist consensus won't survive: The case of classical Mendelian genetics. *Proceedings of the 1990 Biennial Meeting of the Philosophy of Science Association,* edited by A. Fine, M. Forbes, and L. Wessels, 125–139. East Lansing, MI: Philosophy of Science Association.

——. 1994. Genes made molecular. *Philosophy of Science* 61: 163–185.

——. 2004. What was classical genetics? *Studies in History and Philosophy of Science* 35: 783–809.

——. 2008. Beyond theoretical reduction and layer-cake antireduction: How DNA retooled genetics and transformed biological practice. In *The Oxford handbook of philosophy of biology*, edited by M. Ruse, 238–262. Oxford: Oxford University Press.

Watson, J. D., and F. H. Crick. 1953. Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171: 964–967.

Weber, M. 2005. *Philosophy of experimental biology.* Cambridge: Cambridge University Press.