Sequence Matters: Genomic Research and the Gene Concept

Laura Perini[†‡]

Analysis of two key ways of characterizing genes—as causes of phenotypic effects and as genomic DNA sequences—has yielded widespread pessimism that they can be united in a coherent gene concept. This raises important questions about the epistemology of genomic research: If analysis of a genome sequence cannot yield information about genes defined both in terms of their products and their DNA sequence, then what could we learn from it? I investigate basic tools of genomic analysis, argue that they do not reflect the application of either gene concept, and clarify how we learn from genomic research.

1. Introduction. Analysis of gene concepts that capture past and present uses of the term has produced widespread skepticism about the prospects for uniting two key ways of characterizing genes—as causes of phenotypic effects and as genomic DNA sequences—in a coherent concept of the gene.[1] The various ways DNA is involved in the processes that ultimately result in the most proximal effects of gene expression, the protein and RNA molecules often called "gene products," seem to prohibit characterizing genes as entities with both a specific molecular product and a determinate physical identity as a segment of chromosomal DNA whose sequence corresponds to that of the gene product (Fogle 2000; Moss 2003; Downes 2004; Stotz, Bostanci, and Griffiths 2006). If there is no viable gene concept that combines both of these elements, then the research

†To contact the author, please write to: Department of Philosophy, Pomona College, 551 N. College Ave., Claremont, CA 91711; e-mail: Laura.perini@pomona.edu.

‡I wish to thank Cheryl Kerfeld, André Cavalcanti, and James Lennox for their help with this project.

1. Effects can be characterized in various ways, including as a molecular product or contribution to a phenotype or as a difference in phenotype (related to a difference in genetic material).

Philosophy of Science, 78 (December 2011) pp. 752–762. 0031-8248/2011/7805-0004$10.00
Copyright 2011 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/662565 Published online by Cambridge University Press

methods of genomics become somewhat mysterious. If analysis of a genome sequence cannot yield information about genes that are both defined in terms of their gene products and reflected in the DNA sequence—because there is no such thing—then what could we learn from it?[2]

I will call the view that any gene concept incorporating both structural and functional specification of what a gene is cannot be coherently applied in genomic analysis the exclusivity thesis. In a recent article, Piotrowska applies the exclusivity thesis in her assessment of the results of comparative genomics. She starts by asking whether comparative genomics employs a gene concept that "maps onto a *structural* or a *functional* conception of the gene" (2009, 245).[3] She claims that "[if] the gene employed as the unit of measurement corresponds to a structural gene concept, then comparative genomicists must view genes as ordered sequences of DNA," and the comparative claims that result are merely claims about shared nucleic acid sequence (845). If, however, the gene concept involved is a functional concept, then "comparative genomicists must view genes as entities which give rise to phenotypes" and could produce substantive claims about similarity between the compared species (845). However, the analysis applied to genomic data involves searching sequences and evaluating sequences for similarity. Piotrowska's conclusion is that the results are not substantive; in other words, the comparisons are based on a structural gene concept rather than a functional one, so they do not have functional significance and so do not genuinely capture genetic similarity.

Piotrowska has made an important move in extending philosophical analysis of genes to the methods biologists use in research on genomes. Although philosophers of biology have made use of recent results from genomics, they have had surprisingly little to say about how genomes are analyzed, even in works organized around developments in genomics (Moss 2006; Barnes and Dupré 2008). Furthermore, while Piotrowska focuses on the most famous claims of comparative genomics, those reporting the percent of genes shared between two species, there is nothing in her analysis of the bioinformatic methods used in those studies that restricts her conclusion to these particular results. If Piotrowska is correct, it is unclear how the analytical tools applied to genomic sequences could yield any information of functional significance. But the sequence of a genome is of no use to biologists in the absence of the analytical tools

2. A genome sequence is the sequence of nucleotide bases of a full single copy (haploid) set of an organism's chromosomes. Throughout the article, I will refer to genomes as DNA sequences; for simplicity, I do not discuss RNA viruses.

3. She considers no other alternatives, so she is also reasoning from an exhaustivity thesis: genomic analysis must involve application of one or the other type of gene concept. I will address this in sec. 4.

https://doi.org/10.1086/662565 Published online by Cambridge University Press

used to study it; it is what biologists might learn from that sequence that matters. Given the resources, in funding and personnel, now devoted to sequencing and analyzing genomes, Piotrowska's paper raises a high-stakes question about the epistemology of genomic sequence analysis.[4]

Since the problem of understanding what could be learned from analyzing genomic sequences depends on the exclusivity thesis, I will first review the biological phenomena that support the claim that genes cannot be entities specified both as genomic DNA segments and in terms of determinate molecular products. In the next section, I will explain some of the most basic techniques used in analyzing genomes. This will set the stage, first, for investigating whether the analytical techniques typically used in genomics depend on a particular gene concept and, second, to allow for clarification of the epistemic value of genomic research, given that this cannot depend on a gene concept that unifies structure and function.

2. The Gene Concept in Question. The philosophical literature on genes is too extensive to summarize here.[5] Suffice it to say that biologists have used the term 'gene' in so many different ways that, in response, philosophers have offered several different characterizations of types of gene concepts in order to make sense of the variety.[6] A survey of this variety is unnecessary for the purposes at hand since the epistemological questions raised about the study of genomic sequences arise from the two gene concepts that are most prominently involved in that research. I will first explain the biological phenomena that make it difficult to articulate a gene concept that combines these two—that could capture a physical specification of a gene as a nucleotide sequence and as something that is the cause of a functional molecular product. While the production of RNA molecules that play important functional roles in cellular processes is critical, in this article I will focus on the genes involved in protein production.

Two basic processes are necessary for protein production: transcription and translation. Transcription is the production of an RNA molecule from

4. In this article, I do not aim to address the specific kinds of research results that Piotrowska is concerned with (genetic comparisons) but the issue her work raises about genomic research in general.

5. See Griffiths and Stotz (2007) for a helpful starting point.

6. For example, Burian distinguishes between a schematic gene concept, which is connected to a phenotype but without specification of a physical substrate, and a discontinuous gene concept, which does have such a specification (2005). Downes sorts gene concepts into two main categories: narrow gene concepts, which identify genes in terms of DNA sequence, and wider concepts that disperse the referent of 'gene' to other aspects of cell machinery; he also suggests a third, pragmatic concept (2004).

https://doi.org/10.1086/662565 Published online by Cambridge University Press

a segment of DNA, with the DNA providing a template for the sequence of nucleotides in the RNA. Translation is the production of an amino acid sequence from messenger RNA (mRNA); some transcribed RNA play different kinds of roles in the cell from mRNA, like ribosomal RNA (rRNA), and some RNA transcripts are called pre-mRNA because their sequences are altered to produce an mRNA that is then translated. In translation, the sequence of nucleotides in the mRNA directs the addition of amino acids, with three-nucleotide triples (codons) being used to determine which amino acid is added. The region of the mRNA that is 'read' in this way begins with a methionine codon that functions as a start signal, followed by the rest of the codons for the resulting amino acid chain, followed by a stop codon.

In prokaryotic organisms (bacteria and archaea), the RNA initially transcribed undergoes little processing before translation, and there is often a direct correspondence between the sequence of amino acids in the protein product and nucleotides in the open reading frame (ORF), the series of nucleotides from the start to the stop codon. However, this simple model is not an adequate account of gene expression in eukaryotes (animals, fungi, plants, and protists), which have significant post-transcriptional RNA processing. The result of those modifications is that the sequence of nucleotides in an ORF in the genome does not match the sequence of nucleotides in the mRNA that is involved in translation, and thus there is not a direct correspondence between the amino acid sequence of the protein product and the nucleotide sequence in the genome. Even in prokaryotes this correspondence may not hold, due to events like frame-shifting, which occurs during the translation of some mRNAs: rather than 'reading' the nucleotide sequence as continuous, contiguous triples, the reading frame shifts one or more nucleotides during translation. In eukaryotes, the initial RNA transcript may go through extensive processing, typically cutting out some regions (introns) and splicing together others (exons—the parts of the original transcript that is expressed). Sometimes the RNA nucleotide sequence is simply modified, or edited, by changing a base from one nucleotide to another or adding or deleting bases. From the perspective of the resulting protein, there is no segment of genomic DNA with "physical integrity" (Fogle 2000, 4) that corresponds to the amino acid sequence.

Furthermore, some initial RNA transcripts are spliced in alternative ways, depending on the circumstances, creating mRNAs with different subsets of the exons contained in the initial transcript. And there may be multiple initial transcripts produced from the same segment of the genome, by transcribing to capture different reading frames or by transcribing a segment with an ORF that is within a larger ORF. Thus, looking at things

https://doi.org/10.1086/662565 Published online by Cambridge University Press

from the direction of a part of the genome, there is no one molecular product that can be associated with that segment as its genetic effect.

These implications have been stressed repeatedly by authors investigating the gene concept (Moss 2003; Downes 2004; Stotz et al. 2006). Fogle claims that attempts to apply a consensus molecular gene that combines structural and functional features has the effect of "forcing arbitrary choices about the physical edges or the physiological properties of the gene" (2000, 19). This supports part of the assumption that Piotrowska makes: if genome analysis involves applying a gene concept in a nonarbitrary way, that concept involves either structural or functional specification, but not both. I will assume that these are indeed exclusive options. In the next section, I outline some basic analytical tools involved in research on genomes, setting the stage to assess whether that does indeed involve application of a gene concept and, if so, what sort of concept is involved.

3. Analyzing Genomes. There are multiple ways to analyze and compare genomes; in this section I focus on techniques that both are common and are likely candidates for involving gene concepts, in contrast to techniques like aligning whole genomes or determining and comparing guanine-cytosine content.

3.1. Gene Prediction. One very basic part of genome analysis is what the researchers involved call gene prediction: the attempt to identify segments of DNA in the genome that are transcribed and yield a functional product—either RNA or, after translation, protein. In this discussion, I will focus on prediction of genes involved in protein production. As sketched above, this is a multistep process in which DNA is first transcribed to a single-stranded nucleotide sequence, which is an RNA copy of one strand of the template DNA. In translation, the nucleotide sequence of an mRNA directs production of a chain of amino acids. How do you decide which parts of the genome encode proteins? The first step is to search the DNA sequence as a sequence of codons. There are six different reading frames to consider: for each strand of DNA, at an arbitrary starting point, a triplet could start at base pair 1, 2, or 3 and is always read in a 5′ to 3′ direction; the other strand would be read in its 5′ to 3′ direction, which is opposite the first strand, starting at any of the three bases. So the first step is to determine the codons corresponding to the DNA sequence, in each of the six different reading frames, and then analyze the six sequences of codons. In translation, AUG (the codon for methionine) functions as a start signal, and three specific codons function as stop signals. A DNA sequence that corresponds to a codon sequence bounded by start and stop codons is called an ORF. Searching for ORFs

https://doi.org/10.1086/662565 Published online by Cambridge University Press

provides a way to predict protein-encoding regions of the genome, but it is certainly not sufficient to prove that a particular segment of DNA encodes a protein product. Because methionine codons are expected to occur by chance at a fairly high rate in noncoding sequences, long ORFs are more likely to encode proteins (Higgs and Attwood 2005, 286). In eukaryotes, discriminating between coding and noncoding ORFs is more complicated because their proteins are usually made from RNA that has been modified after transcription from the genomic sequence.

Before addressing those complications, however, I wish to emphasize several aspects of the basic technique of searching a genome for ORFs. First, the technique is a straightforward analysis of the nucleotide sequence of the genome. Second, the features that are searched for have known significance (the start and stop signals for translation). This kind of search is using information about the causal role that these specific base triplets play in the process of translation. Third, this method is a discovery tool; it provides a way to identify promising candidates out of the entire genomic sequence but does not guarantee that all the ORFs found encode proteins. When likely protein-encoding ORFs have been located, the sequence of amino acids of the protein can be predicted by using the standard genetic code.[7] The upshot is a prediction that a particular segment of the genome encodes a particular amino acid sequence. This very simple form of analysis thus exhibits features we will see repeatedly. Genome analysis involves information about biological systems. More specifically, it involves knowledge of basic biological processes and how nucleotide sequences are involved in those processes, such as protein translation beginning with a methionine codon. And epistemically, the techniques are best understood as discovery tools.

Often information about the genes, or genome, of the organism is used in gene prediction. In E. coli, upstream features that play functional roles in regulating transcription are highly conserved (Mount 2001, 348). Thus, searching for similar sequences provides a way to predict new genes because similar sequences have a high probability of playing a similar role in regulating transcription. This inference is not based only on a correlation of finding particular conserved sequences upstream of known coding regions; the initiation of transcription involves binding of the transcription complex, which in turn is affected by binding of activator and repressor proteins. Such protein-DNA binding events depend on both the structure of the protein involved as well as the structure—shape, distribution of charge, and so on—of the DNA sequence. That is why the specific order of base pairs matters. The results of such a search are not

7. Since there is some deviation from the standard code, this does not guarantee correct prediction of the amino acid sequence.

https://doi.org/10.1086/662565 Published online by Cambridge University Press

guaranteed to be genes: false positives are possible. Additional analysis, for example, for examining the putative coding sequence for similarity to known protein domains, can increase confidence; some of those methods will be described below. False negatives are also possible; their likelihood depends in part on the set of sequences used to identify conserved features.

Gene prediction is more complicated in eukaryotes. Eukaryotic mRNA is processed in various ways, including removing the noncoding introns and splicing exons together. Because introns are interspersed among the exons, the nucleotide sequence of that part of the genome does not match the sequence of nucleotides in the processed mRNA, so it does not correspond to the sequence of amino acids of the protein that is produced. Mount reports on methods to help distinguish whether an ORF encodes a protein (2001, 343). They all depend on knowledge of protein-encoding sequences, but some of these involve general knowledge about protein production, rather than information about specific known protein sequences; I focus on techniques involving general knowledge here and address the analytical use of known protein sequences in the next subsection. First, in a protein-encoding sequence, the frequency of identical bases occurring three bases apart is higher than in noncoding regions, simply because codon use is nonrandom. Another known feature of protein-encoding regions is that codon use varies among organisms, so information about a particular organism can help identify its genes. Thus, statistical analysis of nucleotide sequences can be used to discriminate between coding and noncoding regions.

One common technique involves the use of hidden Markov models, which are built to analyze sequences in terms of significant types of parts of the sequence: regulatory sequence, start codon, exons, splice sites, introns, and stop codon. These models include probability assignments for each nucleotide in a given state and for transition from one state to another. They can be used to identify the most probable series of state transitions and thus the boundaries of exons, introns, and so on (Eddy 2004). Mount claims that methods like GLIMMER (an interpolated Markov model) are reliable ways to find genes that are not similar to previously identified proteins, in the sense of coding for similar sequences of amino acids (2001, 351). This means that DNA sequences that have rather abstract kinds of similarities in terms of nucleotide sequence (e.g., similarities at splice sites rather than overall nucleotide sequence similarity) can be used to predict the presence of protein-encoding sequences when there is no similarity between the two at the level of an amino acid sequence.

3.2. Sequence Alignment. The other approach to identifying protein-encoding regions of the genome, using specific information about known proteins (or protein-encoding sequences), is one of the most basic tech-

https://doi.org/10.1086/662565 Published online by Cambridge University Press

niques of genomic analysis: sequence alignment. DNA sequence can be aligned with other DNA, protein sequences can be aligned with each other, and translated nucleotide sequences can be aligned with proteins and with each other. For example, BLAST (Basic Local Alignment Search Tool) is available in several different versions, so that both nucleotide and protein sequences can be used to query both nucleotide and protein databases (National Center for Biotechnology Information 2011). Alignment tools are designed to compare sequences—by using either individual sequences specifically chosen for alignment or a sequence to search a database to find similar sequences. These tools are widely used in analyzing genomes. For example, gene prediction can be followed by using a detected ORF as a query for a protein database: if the translation of the ORF is an amino acid sequence appropriately similar to a known protein, then this increases the likelihood that the ORF encodes a protein.

Alignment programs compare two or more sequences, looking for a series of bases (or amino acids) in the same order in the compared sequences. The programs compute the best way to line up the two sequences on the basis of different scoring for matching residues (which improve the score) versus gaps and substitutions (which detract). Protein alignment programs make use of amino acid substitution matrices based on data from known protein sequences, enabling differential scoring of specific amino acid matches and mismatches: for example, the BLOSUM62 tryptophan-tryptophan match does not score the same as a leucine-leucine match (Mount 2001, 88). The output of a BLAST search of a database includes both the aligned sequences and, for each alignment, an E value, which is the number of sequences in the database expected to give as good an alignment score to the same length query, by chance alone. A low E value, thus, is a good indicator that the alignment is significant, and the sequences are related (Higgs and Attwood 2005, 139).

Homology in genomics refers to genes that share a specific kind of evolutionary history: homologous sequences descend from a common ancestor sequence (Higgs and Attwood 2005, 9). The details of the descent relationship can vary in significant ways; researchers are often interested in learning whether two homologous sequences are orthologs (i.e., the origin of the two sequences occurs with the splitting of lineages) or whether the original source of the two sequences happened through a duplication, in which case they are paralogs. The difference between homology and similarity is significant biologically. Analytical methods are directed at detecting homology, and they are designed to identify similarities that provide evidence for an evolutionary relationship between the sequences, including more detailed information like distinguishing between orthologs and paralogs. Alignment tools have a variety of uses beyond gene prediction because they are built to detect these evolutionary relationships.

https://doi.org/10.1086/662565 Published online by Cambridge University Press

A known protein-encoding sequence can be used as a query to search a genome for orthologous sequences; a genome can be searched against itself to find paralogs.

Alignment programs allow researchers to identify alignments that are likely (although not guaranteed) to reflect an evolutionary relationship. Evolutionary relationships between sequences, in turn, are relevant to function: for example, closely related orthologs are likely to function in the same way, contributing to the production of similar proteins under similar circumstances. The fallibility of alignment programs motivates ongoing development and analysis of these programs (Lunter et al. 2008). Like gene prediction, alignment programs should be understood as discovery tools.

4. Genomic Analysis and the Gene Concept. The question raised by philosophical analysis of the gene concept is, if "genes" are not objects defined by both genomic sequence and molecular product, then how could analysis of genomic sequences produce results—purportedly about genes—that mean anything at all? Piotrowska's response to this problem is that comparative genomics cannot produce substantive—that is, functionally relevant—results. Her reasoning depends not only on the exclusivity thesis but also on an assumption of exhaustivity: she assumes that genomics must involve either a structural or a functional concept of the gene. Since the researchers in this area use the term "gene" frequently, this seems initially plausible. However, on closer investigation it is not clear that a gene concept is deeply involved in genomic research. The techniques used to analyze genomic sequences are not driven by a theory of the nature of the gene, and in particular, they are not driven by a structural concept of the gene. They draw on a variety of kinds of information about how DNA sequences are involved in the processes of transcription and translation and known statistical differences between the composition of coding and noncoding regions, background knowledge of the biochemistry involved in protein-DNA binding, and evolutionary theory. But use of this information does not imply that these analytical techniques are being used to investigate a kind of a thing that is defined as a segment of the genome. For example, aligning genomic sequences can be used to find orthologous segments of the genome, so these techniques make it possible to use a sequence from one organisms' genome, known to be transcribed and spiced in alternate ways, to search for an ortholog in another organism. Because alignments can provide evidence of an evolutionary relationship between the two sequences, and that in turn provides some evidence that the two DNA sequences are involved in cellular processes, like transcription, in similar ways, this procedure offers a way to find other cases of

https://doi.org/10.1086/662565 Published online by Cambridge University Press

alternative splicing.[8] These tools are not designed around a concept of the gene with "physical integrity." The gene prediction tools may not be as successful in identifying disparate genomic DNA sequences that contribute to the production of a protein; they may be more prone to false negatives in such cases. The fact that the tools are less reliable in such cases does not show that their use involves a structural gene concept.

Similarly, the fact that the researchers are using these techniques to identify and compare putative molecular products does not imply that their results must be understood in terms of a functional gene concept, that is, finding and comparing genes defined as causes of molecular products. The analytical tools used in functional genomics can also be used to find and compare noncoding genomic sequences, which suggests the results of applying these tools should not be characterized as finding and comparing 'genes' defined in terms of molecular products.

If genomic research is not aimed at producing knowledge about the gene—defined either functionally or structurally—then what is its epistemic value? Genomic analysis allows for identifying causal relationships between parts of genomes and molecular products and identifying different kinds of evolutionary relationships among parts of genomes (such as orthologs and paralogs). These in turn have implications for whether the homologous genes share function. In general, the results require further support; these are discovery tools providing some evidence but not conclusive support. The epistemic importance of discovery in this context should not be underestimated: genome sequences can have billions of base pairs. There are hundreds of sequenced genomes, with more in the pipeline (see, e.g., the projects completed and in development at the Joint Genome Institute [2011]). It is only through the use of search tools that any hypotheses at all can be formulated from genomic sequence data.

5. Conclusion. This study is a preliminary effort to clarify the epistemology of genomic analysis. However, the survey of the most basic techniques involved has had several results. It shows that this kind of research is not constrained by a particular gene concept, nor are the results used to fill in or support a particular theory of the nature of the gene. Rather, various kinds of information about the relationships between genome sequences and protein production are used to find new cases. That information can involve stereotypical gene features, like the ORF structure, as well as information from specific cases (as when a known protein

8. This is not the most efficient way to find cases of alternative splicing; if that is a goal, as opposed to finding orthologs of a specific case, then analysis of a sequenced cDNAs (complementary DNAs) or ESTs (expressed sequence tags), or, even better, the transcriptome, is required (see Filichkin et al. 2010).

https://doi.org/10.1086/662565 Published online by Cambridge University Press

sequence is used as a query against a new genome). Higgs and Attwood characterize genomics as a "knowledge-based" discipline, to highlight the fact that it is not guided by some overarching theory but rather depends on many different kinds of facts known about organisms and evolution (2005). This study validates that characterization. Finally, identifying the reasoning involved in genome analysis as a discovery process both clarifies its epistemic value and sets the stage for further philosophical investigation.

REFERENCES

Barnes, Barry, and John Dupré. 2008. *Genomes and What to Make of Them*. Chicago: University of Chicago Press.

Burian, Richard M. 2005. "Too Many Kinds of Genes? Some Problems Posed by Discontinuities in Gene Concepts and the Continuity of the Genetic Material (1995)." In *The Epistemology of Development, Evolution, and Genetics: Selected Essays*, ed. Richard Burian, 166–78. New York: Cambridge University Press.

Downes, Stephen. 2004. "Alternative Splicing, the Gene Concept, and Evolution." *History and Philosophy of the Life Sciences* 26:91–104.

Eddy, Sean R. 2004. "What Is a Hidden Markov Model?" *Nature Biotechnology* 22:1315–16.

Filichkin, Sergei A., H. D. Priest, S. A. Givan, R. Shen, D. W. Bryant, S. Fox, W. K. Wong, and T. C. Mochler. 2010. "Genome-Wide Mapping of Alternative Splicing in *Arabidopsis Thaliana*." *Genome Research* 20:45–58.

Fogle, Thomas. 2000. "The Dissolution of Protein Coding Genes in Molecular Biology." In *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*, ed. Peter Beurton, Raphael Falk, and Hans-Jörg Rheinberger, 3–25. New York: Cambridge University Press.

Griffiths, Paul E., and Karola Stotz. 2007. "Gene." In *Cambridge Companion to the Philosophy of Biology*, ed. David L. Hull and Micael Ruse, 85–102. New York: Cambridge University Press.

Higgs, Paul G., and Teresa K. Attwood. 2005. *Bioinformatics and Molecular Evolution*. Malden, MA: Blackwell.

Joint Genome Institute. 2011. Genome projects, U.S. Department of Energy, Washington, DC, http://www.jgi.doe.gov/genome-projects.

Lunter, Gerton, A. Rocco, N. Mimouni, A. Heger, A. Caldeira, and J. Hein. 2008. "Uncertainty in Homology Inferences: Assessing and Improving Genomic Sequence Alignment." *Genome Research* 18:298–309.

Moss, Lenny. 2003. *What Genes Can't Do*. Cambridge, MA: MIT Press.

———. 2006. "Redundancy, Plasticity, and Detachment: The Implications of Comparative Genomics for Evolutionary Thinking." *Philosophy of Science* 73:930–46.

Mount, David W. 2001. *Bioinformatics: Sequence and Genome Analysis*. New York: Cold Spring Harbor Laboratory.

National Center for Biotechnology Information. 2011. BLAST (Basic Local Alignment Search Tool), U.S. National Library of Medicine, Bethesda, MD, http://blast.ncbi.nlm.nih.gov/Blast.cgi.

Piotrowska, Monika. 2009. "What Does It Mean to Be 75% Pumpkin? The Units of Comparative Genomics." *Philosophy of Science* 76:838–50.

Stotz, Karola, Adam Bostanci, and Paul E. Griffiths. 2006. "Tracking the Shift to 'Post-genomics." *Community Genetics* 9:190–96.

https://doi.org/10.1086/662565 Published online by Cambridge University Press