Understanding Contemporary Genomics

John Dupré
ESRC Centre for Genomics in Society
University of Exeter

Recent molecular biology has seen the development of genomics as a successor to traditional genetics. This paper offers an overview of the structure, epistemology, and (very briefly) history of contemporary genomics. A particular focus is on the question to what extent the genome contains, or is composed of, anything that corresponds to traditional conceptions of genes. It is concluded that the only interpretation of genes that has much contemporary scientific relevance is what is described as the "developmental defect" gene concept. However, developmental defect genes typically only correspond to general areas of the genome and not to precise chemical structures (nucleotide sequences). The parts of the genome to be identified for an account of the processes of normal development are highly diverse, little correlated with traditional genes, and act in ways that are highly dependent on the cellular and higher level environment. Despite its historical development out of genetics, genomics represents a radically different kind of scientific project.

An ancestor of this paper was written for a symposium on "Proof and Demonstration in Science and Mathematics." This presented an immediate difficulty that I was unsure whether there was anything to be said on these topics relating to my current areas of study, genetics and genomics.

A version of this paper was presented at the Athens-Pittsburgh Symposium in the History and Philosophy of Science and Technology in Delphi in June 2003, and benefited from the comments of several members of the audience there. I am especially grateful for detailed comments by Richard Burian, who also helped me to remove some of the more egregious falsehoods from my Lakatosian history. The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. This work was part of the program of the ESRC Research Centre for Genomics in Society (Egenis). My thoughts on these topics owe much to discussions with my colleagues in Egenis. I am especially indebted to Steve Hughes for continuing education about contemporary genomics. I also thank an anonymous referee for pointing out some factual errors.

Perspectives on Science 2004, vol. 12, no. 3

Proof, at any rate, is not a concept I often encounter anywhere in biology. Certainly there is plenty of evidence for some biological claims, but I'm not sure these generally amount to anything that would count as proof. Proof in mathematics is a much more familiar idea, but in so far as it points to a logical relation between axioms and theorems, its application to contemporary science suggests an antiquated philosophy of science which, at least in the context of biology, has been almost entirely discredited. There is, of course, a tradition in the philosophy of science of thinking of explanations as derivation from laws and initial conditions, which raises obvious parallels with deduction from axioms. But nowadays philosophers of biology are much more inclined to talk about models than about laws, and are generally quite skeptical even of the existence of biological laws (see, e.g., Lloyd 1994).

A quite different connection might be through the use of biological evidence in juridical contexts. It is often claimed for instance that the analysis of DNA found at crime scenes provides proof of guilt or innocence. Interesting though such contexts undoubtedly are, and interesting though the issues they raise in understanding juridical proof may be, I doubt whether they raise profound philosophical issues in the study of science.

Demonstration is, I think, a broader concept, and perhaps holds more promise. It can, of course, be more or less synonymous with proof, as again in the context of mathematics and as illustrated by the initials QED which schoolchildren used to be required to add to the conclusion of what they fondly hoped were mathematical proofs. In simpler times, demonstrations were an important part of science pedagogy. A physics teacher might hang weights of various sizes on a spring, measure the length of the spring, and plot the length of the spring against the numbers stamped on the shiny brass weights. Subject only to our confidence that these numbers corresponded to a real property of the weights, something rendered intuitively plausible by their visible sizes—the one marked 5kg looks quite a few times as big as the one marked 1kg, for instance—this might reasonably be taken as a demonstration of Hooke's law. If one attends carefully to the performance, the demonstration, one can more or less see that Hooke's law is true (or for the very skeptical, true here, today, anyhow).

In striking contrast to this simple demonstration, I recently had the good fortune to receive a tour of the Sanger Centre, outside Cambridge, where a large part of the Human Genome Project, the sequencing (more or less) of the entirety of the genetic material in a human cell, was carried out. Over the reception desk an electronic display flashes a stream of Cs Gs As and Ts which, we are informed, constitute a real time read out of some

DNA that is being sequenced somewhere on the premises. Touring the building where the sequencing actually takes place, the first stop is a room in which large robots stick tiny probes into Petri dishes and then into rectangular arrays of test tubes. Spots on the nutrient gel in the Petri dishes, we are told, contain bacteria infected by viruses with pieces of human DNA. We next peer through a window in the door of a room containing small but expensive machines that perform the polymerase chain reaction, the process that multiplies the quantities of DNA generated in the first process to the quantities required for the sequencing machines. These latter, finally, occupy a warehouse-sized space in which conversation is rendered difficult by the hum of the powerful cooling systems. There are perhaps a hundred of these machines, each connected to a familiar looking desktop computer, all busily sequencing genomes. The room is largely devoid of human activity, except for the occasional lab assistant carrying trays of material to be fed into the machines. A separate building, which I did not see inside, houses the bioinformatics operation, in which the output of all these machines, and others like them around the world, are chewed over by powerful computers.

My hour or two touring the sequencing centre might perhaps be referred to as a demonstration of a state-of-the-art genomics laboratory. But it is very clearly far removed from the simple demonstration of Hooke's law. By contrast with my modest faith that the weights have been accurately labeled, in the Sanger Centre everything is taken on trust. If the entire operation was a mock-up by Lucasfilms, I'm sure I would be none the wiser.

Perhaps the most interesting moral of this comparison is the way in which it points to the division of labor in much of modern science. Though there are of course plenty of biologists who understand the basic biological principles underlying the various bits of machinery in the sequencing lab, it's a fair bet that few or none of them know in any detail how all of these machines work. Moreover, even those who know how they work in some detail surely don't normally have the expertise in operating them possessed by experienced technicians familiar with their quirks and occasional malfunctions. And even those technicians surely don't have the expertise of the engineers who design and construct the machines or who repair them when they malfunction in serious ways.

To cut a long story short, a project such as the sequencing of the human genome involves the collaboration of thousands of people with hundreds or thousands of different forms of expertise, not to mention requiring many years of work by this large and diverse group of people. Clearly no one could offer a demonstration that the human genome was. . . . One is

reminded of Descartes' concern that a proof should be compact enough to be held in the mind at the same time, though presumably Descartes never dreamed of anything quite this far from meeting this optimistic ideal. If one has confidence that the published drafts of the human genome bear some close relation to something in reality this is based not on proof or demonstration, but on trust. And this is as true for Sir John Sulston or Craig Venter as it is for the casual reader of *Nature* or tourist in the Sanger Centre.

No doubt this is all too ambitious. Surely within the practice of genetics and genomics, as within any human practice with even a minimal intellectual content, there are arguments. For instance:

> This gene codes for the *Bacillus thuringiensis* toxin
> If we insert it into the genome of this plant, the plant will produce BT toxin
> BT toxin poisons insect pests
> Therefore, if we insert this gene into this plant, the plant will poison insect pests.

This argument is plausible, if a bit enthymematic. One premise that might start to flesh it out is:

> If we insert a gene for x into a (living) genome then that genome will produce x.

This premise shows us that the argument, whether or not plausible, is not sound. For the missing premise is certainly false. There are lots of reasons for this falsity. One of the most interesting involves the familiar redundancy of the genetic code. Amino acids, the constituents of proteins, are coded for by as many as six different base-pair triplets. However, different organisms tend to use different triplets preferentially and will be disproportionately equipped with extra-nuclear equipment for reading the preferred codons (Ikemura 1981). Consequently they may be very bad at transcribing a gene from a distantly related organism. More simply, whether a sequence is transcribed will depend very much on where it ends up in the genome, on its spatial relations to other genes, especially promoter and suppressor sequences, and even to other structures in the cell. Current techniques for inserting genes into alien genomes are thoroughly hit or miss as far as where the genes end up.

Another reason that the gene may fail to produce the toxin is that the plant may die before it has a chance to do so. If the inserted gene should land in the middle of a sequence of the genome vital for the plant's func-

tioning then the plant will not function. Inserted genetic material may also have a range of effects on the host organism distinct from those intended (pleiotropy), and these may be harmful or fatal.

The relevant moral of these genomic factoids is that genomic events are diverse and specific. One familiar model of scientific argument, that most closely connected to mathematical ideas of proof and demonstration, essentially involves generalizations—traditionally thought of as scientific laws—and generalization is a risky business in biology generally and genetics in particular. The simple example just discussed illustrates the difficulty. The attempt to convert such simple generalizations into exceptionless laws would be extremely difficult if not impossible. Such considerations lead naturally to the conclusion that there are few if any laws that apply to genes.

And there is an even more basic reason for the lawlessness of genes: as I shall explain shortly, it is doubtful whether there are any genes at all. My point so far is not, of course, that no one engaged in genetics or genomics ever deploys any kind of argument. My thesis is rather that argument has no special role in genetics beyond that which it plays in any other intelligent human activity. There are no general patterns of argument to be found, certainly no premises that recur across indefinitely many different genetic arguments.

One of the reasons for the lack of such recurring premises is that genes, the apparent subject matter of genetics, if they should be said to exist at all, are highly diverse entities and not the kinds of things that might be the subjects of broad generalizations. We can, perhaps refer relatively unproblematically to whole genomes. A biologist colleague likes to define the genome as "a space in which genetic things happen."[1] Genes are then, perhaps, the things that things happen to in genomes. But all kinds of different things happen in genomes, and they happen to different kinds of things. Generality in genetics and genomics applies to some interesting extent to the tools, techniques, and instruments that can be used to provide insight into genetic events. But genetic events themselves are hardly more homogeneous than, say, things seen through a telescope, or a microscope.

A more positive way of stating the point is the following. Traditional philosophy of science sees central concepts as opening up the possibility of discovering laws of nature or, at any rate, general knowledge of nature. The example of genes suggests something quite different: the function of this concept is rather to allow us to talk about lots of different things (see

1. Thanks to Steve Hughes for this illuminating idea.

Rheinberger 2000 for a related account of the term "gene"). Such a concept facilitates communication between people with different but related concerns, and facilitates continuity between successive historical inquiries. It may also provide a risk of serious misunderstanding. This risk is probably minimal in the case of working scientists communicating their results to one another, as far more specific, local interpretations of a word such as "gene" will be expected and provided. Misunderstanding arises rather as scientific results disseminate to different areas either of science or to other domains of human life, and such dangers may be exacerbated by obsolete philosophy of science. At any rate, there is a likely role for philosophers of science in attempting to delineate the diversity of meanings of such complex concepts and the fact that such meanings are diverse.

Are there genes?

Understanding the problems with the concept of a gene requires a brief excursion into scientific history, though I'm afraid the present excursion will be somewhat Lakatosian in character.[2] We might begin the Lakatosian enterprise by imagining that Mendel invented the term gene (the footnote attributes this to Willhelm Johannsen, in 1909). At any rate, the tradition of transmission genetics generally supposed to have been inspired by Mendel's work, and epitomized by the famous *Drosophila* experiments of Morgan and Müller, was concerned with genes as hypothetical factors responsible for differences in phenotypes. The gene for red eyes was whatever caused some flies but not others to have red eyes. Of course, nobody supposed that this was the complete cause, as if the gene was something that you could dump in the laboratory disposal bin, and the bin would grow red eyes. But it was the factor that caused the difference in the developmental process that led to the animal having its distinctive eye color.

Inevitably this program inspired an interest in the question, what (if anything) is the physical instantiation of these hypothetical factors. Attention quite quickly focused on chromosomes as the likely location for genes, and in 1927 Müller provided evidence for this hypothesis by establishing that x-ray damage to chromosomes could produce genetic changes in flies. In 1944 Oswald Avery argued that the physical basis of heredity was DNA on the basis of experiments in which DNA was transferred from

2. Lakatos (1980) famously suggested that history of science should mainly be concerned with "rational reconstruction" of what should have happened to best explain the current state of our knowledge, with the actual history confined to the footnotes. Perhaps unlike Lakatos, I do not mean to imply any disrespect for the very important business of real history. Unfortunately I am not equipped to provide it.

pathogenic bacteria into a harmless related species, and thereby transmitted the pathogenicity. This, however, remained controversial. In 1953, as we all know, Crick, Watson, and others disclosed the chemical structure of DNA, and the basis of its capacity for replication. This quite quickly established consensus on the identification of DNA as the genetic material. In 1966, finally, the genetic "code" was "cracked," and the basis of the ability of strands of DNA to determine the production of specific polypeptide chains was understood.

In an obvious sense the elucidation of the structure of DNA was the culmination of the project of transmission genetics. But it was also, in a less obvious sense, the beginning of the end of that project—for it initiated the process of seeing that there really weren't any Mendelian genes, or anyhow not many. This was in fact the conclusion of discussion that might perhaps be said to have inaugurated contemporary philosophy of biology, the question whether the Mendelian gene could be reduced to the molecular gene. The argument that it could not was stated in David Hulls' classic introduction to the philosophy of biology in 1974. The already uncontroversial central premise of Hull's argument was that the relations between molecular genes and phenotypic traits were many/many. A typical molecular gene would have a variety of effects on the phenotype, and any phenotypic trait would require numerous molecular genes for its realization. So the characterization of genes in terms of their phenotypic effects seems drastically underdetermined.

This is, to put it mildly, a simplified story. As I already noted, classical geneticists did not typically have crassly naïve views of the role of genes and intended only to refer to differentiating causes (see, e.g., the sophisticated discussion of the developmental relations between gene and trait in Morgan et al. 1915). The point about many/many relations does not show that a molecular gene may not, under normal conditions perhaps, make the difference between one phenotypic condition and another.

But there is a more important point. The initiation of the program of molecular genetics inevitably directed attention towards development: what are the processes by which genes affect the development of the organism. From the point of view of development, the question of genetic difference is only of tangential interest. It's of more fundamental interest to see how the eye develops at all than to know why it should be red, purple, orange, or crimson. And many or most of the genes involved in normal development aren't even candidates for Mendelian genes, because they are required for development, and the only distinguishable phenotypic state connected to variations in such genes is non-viability

The importance of the relation of Mendelian and molecular genes is a concern that the Mendelian tradition might seem to license a general pro-

gram of identifying genes in terms of their phenotypic effects. (This is a version of what Lenny Moss [2002] has described as the preformationist conception of the gene—the gene as carrying the information necessary and sufficient for the production of a particular trait.) Of course, this is the only way that the tradition could possibly identify genes and it would be quite unfair to accuse its exponents of making any such ungrounded universal claim. Still, what we do see is why there is a strong discontinuity between the Mendelian and molecular traditions. This is just that while the tradition gave access to some of the molecular phenomena, and motivated the search for the molecular phenomena, the phenomena that that search ultimately revealed were not even generally the kinds of things that Mendelian genetics had investigated.

So what is a molecular gene? The natural move in the light of the first decade or so of information about the actual function of DNA was to suggest that a gene was a bit of DNA that contained the code for producing a functional polypeptide, or protein. Thus the connection would be maintained with some product for which the gene was responsible, but the product would be identified much nearer to DNA itself in the causal chain. As a matter of fact, it is quite common to hear this conception of the gene defended to this day (for a sophisticated version of this sort of view, see Waters [1994]).

This conception of the gene is, however, of little general use for analyzing the genome. To begin with, possibly as much as 95% of the human genome (proportions vary for different species) doesn't appear to code for anything or even to have any function at all, and this is often referred to as "junk" DNA. Of course having no known function is not the same as having no function, and it remains possible that all kinds of further functions may be discovered. It seems increasingly likely that the three-dimensional structure of the whole genome may be functionally important, in which case some or most "junk" DNA will be functionally relevant to maintaining this structure. If there is genuine junk in the genome, it is an interesting speculation that this is the DNA to which Richard Dawkins' notorious conception of "the selfish gene" may really apply: this is DNA that exists because it has successfully competed for space in the genome. From the point of view of the organism it is a mere parasite. All this is, however, perhaps a rather minor issue. If, to use a standard abusive expression, the genome were composed of genes "like beads on a string," then all the presence of junk DNA would show would be that there turns out be a lot of string and not so many beads.

Even within the 5% of the genome that seems definitely to be functional, only about 60% is both transcribed into RNA and translated into polypeptide chains. There are, in addition, sequences involved in a variety

of ways in promoting, suppressing, terminating, and activating other sequences. So only about 3% of the genome even holds out the hope of fitting the definition under consideration. But even this modest target can quickly be seen to be unreachable. Sequences identified as "genes," it now appears, are typically composed of alternating coding sequences and sequences which, while often functional, are not part of the sequence for which the gene as a whole codes. These are known, respectively, as exons and introns. All or some of the exons, finally may be transcribed and then assembled into a variety of distinct and often functionally different proteins, sometimes employing in addition coding sequences from other parts of the genome.

A final point, the importance of which is increasingly being realized is that there are variably transient, but heritable, changes to the genome, that can have major functional consequences. Most important of these is the process of "methylation" a modification of the cytosine molecule, one of the bases in the DNA sequence, that affects the activity of a particular coding sequence. The importance of this process is currently being explored in the "Epigenome Project," one of the main successors to the human genome project. This phenomenon emphasizes the extent to which DNA is increasingly perceived as interacting with other elements in the cell, and indeed indicates the accelerating demise of the one time "central dogma" of molecular biology that postulated a strictly unidirectional flow of information from DNA to RNA to protein.

Without going too deeply into these complexities, what emerges can also be seen as a recurrence of the many/many problem that derailed reductionist aspirations for the relation between phenotypes and genotypes. Even between DNA sequences and polypeptides there are many/many relations: a DNA sequence may be involved in the production of a variety of polypeptides, and the production of a polypeptide will normally involve a variety of often spatially distinct DNA sequences. One thought—more likely to occur to a philosopher than a biologist, I suspect—is that one might still maintain the principle one polypeptide, one gene (though not vice versa) and simply recognize that genes had proved to be overlapping and spatially discontinuous entities.[3] But even apart from the rather serious objection that this will overturn most or all existing genetic nomenclature, it fails for more technical reasons. The processes of polypeptide assembly do not necessarily end with translation from RNA to amino acid chain. One salient case is that of chains that split into smaller units after translation. Sometimes these units are identical, some-

3. This idea is discussed, and the difficulties explained, in detail by Fogle (2000).

times different. In the latter case it appears that we must find some way of avoiding the conclusion that all the polypeptide fragments are products of the same gene. Any way we find of doing this is likely to force us to say that the identical products in the former case all are products of different genes. The point is just that the diversity of the processes intervening between DNA sequence and functional protein is such as to make it an unpromising venture to look for some uniform relation between the latter and some privileged part of the former held to have a canonical causal responsibility for it. So, it appears that we cannot use the protein products to base a taxonomy of bits of the genome, and the problem of dividing the genome into genes remains unanswered.

To recapitulate: Mendelian genes, postulated causes of differences between conspecific organisms are, at the molecular level, scarce and equivocal. They are scarce not only because a large proportion of the genome does not even contain candidates for Mendelian genes, but also because much of that which does cannot vary in functionally significant ways without fatally derailing the development of the organism. And they are equivocal because genes are pleiotropic, having a range of different effects on the organism. If we think of genes as "made molecular," as the components of the genome, then Mendelian principles are of little use in delimiting genes.

If, on the other hand, we start with the concrete physical genome, we might perhaps think of genes as the functional constituents of the genome. Unfortunately from this perspective, the delineation of genes appears to be massively underdetermined. There are many different kinds of such functional constituents and, moreover, functional constituents themselves have smaller functional constituents. Is an exon a gene? An intron? For that matter, why not a base pair?

If the Mendelian gene concept is largely inapplicable and the molecular gene concept hopelessly indeterminate, it begins to look as if we would do well not to talk about genes at all.

One reason this may sound surprising is that we not only talk about and hear about genes on a daily basis, but we even learn with considerable regularity that scientists have discovered them. We are naturally inclined to attempt to make sense of this talk. I suggest that most of this talk assumes a concept that has not been sufficiently recognized, what might be called the developmental defect concept.[4]

4. An anonymous referee correctly pointed out that the following discussion considers only deleterious germline mutations, whereas the concept of genetic disease also extends to disease caused by somatic cell mutations. It would be possible to insist that somatic cell

It is not uncommon for discussions of behavioral genetics to establish the credentials of their subject by referring to the genetic disorder phenylketoneuria (PKU). This condition involves the inability to metabolize phenylalanine. The accumulation of this amino acid leads, in turn, to various physical problems and a degree of mental retardation. The disorder is caused by any of a range of mutations in both alleles of the sequence that codes for the enzyme phenylalanine hydroxylase. The pathological condition, PKU, is commonly thought of as a monogenic disease. This, in turn, is naturally interpreted as meaning "a disease caused by a single gene." But we can immediately see that this isn't quite right. The disease is caused by the disfunctionality of a particular gene. And the various mutations that lead to such disfunctionality are not in any natural sense genes for PKU but dysfunctional variants in a genetic region that codes for phenylalanine hydroxylase. The referent of the phrase "gene for PKU" therefore is not a physical object at all, but a set of defects in another object, a coding region involved in the production of a particular protein. There is, I suppose, a technical interpretation of the phrase "gene for x" according to which any of these defective regions is a gene for PKU: the defective allele makes a difference to the developing phenotype. No one of them, however, is *the* gene for PKU. But there is not even a technical sense in which the functionally unimpaired version of the gene is a gene for the prevention of PKU. (It does of course function in the production of phenylalanine hydroxylase, but we have already considered the reasons why it would be misleading to call it the gene for phenylalanine hydroxylase.) Compare with this the idea that the heart is an organ for preventing oxygen-deprivation induced brain damage.

It is striking that perhaps the most familiar roughly Mendelian human physiological trait is a nice example of the developmental defect gene concept, namely eye color.[5] Blue eyes are, roughly speaking, the result of a recessive genetic defect in the production of the pigment that gives eyes their proper brown color. Since in Northern latitudes this defect has no serious consequences, such defects have accumulated in some populations to the point where blue eyes have become the norm. But there is no gene for blue eyes in the rather strong sense that the cause of this trait is a pure ab-

---

mutations can be included under a sufficiently broad conception of development. However, given the clear distinction between these cases it is no doubt better to distinguish them, and recognize a wider range of applications of the term "genetic disease." I do not think this correction significantly affects the philosophical argument here (and nor, I am pleased to say, did the referee).

5. As usual, this example is really much more complicated as there is also a gene for green eyes, and not all color variation has been genetically explained.

sence. There is a little more to be said for talking about a gene for brown eyes, though certainly all the standard problems of pleiotropy, polygeny, and so on will make the terminology liable to mislead.

When we refer to a "gene for x" it is natural to suppose that we are referring to a gene the physiological function of which is to produce x. If behavioral genetics is the study of forms of behavior caused by identifiable genes, then PKU is completely irrelevant to the subject. The same is equally true for familiar physiological disorders. We can, of course, insist on using the phrase "gene for x" in a different technical sense derived from Mendelian genetics and also adopted by some evolutionists. The problem then is that behavioral genetics will have only the slimmest connection with the causes of behavior. My own reading of the evidence is that there is, in fact, little reason to expect that genetic differences will be useful in explaining behavioral or mental differences beyond the cases of serious incapacity caused by malfunctioning genes. This would, indeed, make the developmental defect gene concept the appropriate one in this context, but would also undermine most of the publicly expressed pretensions for this field of study. To take one example, it was recently widely reported in the press that the gene for human arts and culture had been found. The consilient evidence for this claim was, first, that a gene had been isolated with a mutation that occurred subsequent to the split of the human lineage from that of the great apes; and, second, that damage to this gene caused people to be deficient in artistic and cultural skills. I hope it is clear that this does not provide the slightest shred of an argument for the discovery of a gene for arts and culture in any normal interpretation of that expression.

Let me offer one more simplistic summary of the simplistic historical narrative. For much of its history genetics was driven by a hypothetical kind that it saw itself as investigating, the gene. As we gradually identified the material referent of this hypothetical kind and were able to learn something about how its instances worked and what they did, it became increasingly clear that they were not a kind at all but a diverse set of molecular objects and processes. There is perhaps a legitimate kind, DNA sequence, and some instances of this kind do indeed do something interesting: they are transcribed into RNA sequences, some of which are translated into proteins.

I won't go into much detail here about positive accounts of the gene. A number of accounts have been offered by people who have come to terms with the sort of complexities just discussed. Most of these, in my view, have the fatal defect of legislating a concept much narrower than historical conceptions of the gene and a concept too closely tied to a particular theo-

retical idea.[6] At the opposite extreme, and rather more promisingly, Hans-Jorg Rheinberger (2000) has suggested that central scientific concepts, like the gene, function precisely by remaining sufficiently vague to allow communication between all the various groups that have an interest in talking about such things, but very diverse accounts of what it is they are talking about.

I do think there must be something right about this view. However it does at least need supplementation to account for the great precision with which particular genes are referred to in narrow scientific contexts. I shall offer a rather different tentative suggestion. There are interesting parallels with a topic I have been interested in for many years, the so-called species problem. It seems almost indisputably impossible to find a definition of the species that is applicable across the whole range of biological diversity. There are partisans for a variety of species concepts, and these supporters take various attitudes to the bits of the biological world they don't adequately cover (they're not important; more research is needed; they don't form species at all; etc.) An inevitable, and in my view correct, reaction to all this is pluralism: there is no definition of species and groups of various kinds should qualify. Most pluralists nevertheless try to hang on to some theoretical core to the concept, generally that a species have some kind of phylogenetic coherence.[7]

My own preference is for total abandonment of such theoretical commitment. One reason for this is the insistence that "species" is not primarily a theoretical concept at all, but a classificatory concept. (Certainly this accords with the principle of priority, which is an important one within scientific taxonomy.) It is naturally assumed that these will coincide, as seems to be the case, for example, with the classification of chemical elements. But the path to pluralism reveals that this is not in fact the case for biological kinds. One is driven to pluralism by the realization that theoretical principles that seem to work nicely for some domains of biological classification turn out to be wholly inapplicable to others, so that attempts to provide a monistic account of what a species is leave us unable to classify large areas of biological diversity. When theoretical conceptions of the species are applied to practical taxonomy, theory can even become an enemy of classification. Changes in theory will lead to change in classification and stability is an obvious desideratum of a classificatory scheme. Taxonomic conservatism must be recognized as an important criterion in

6. For example, the interesting account of Beurton (2000) seems to me to suffer from this defect.

7. My views on this topic are explained in detail in Dupré (2002), chs. 3 and 4.

assigning species names to groups of organisms, and even more so when it is recognized that there is no universally adequate theoretical conception to which classification should be answerable.

Turning now to the naming of genes, I was struck recently by the following sentences on a major bioinformatics website:

> "Keeping stable names for "things," such as genes, in databases is very important. This allows scientists in different labs around the world to be confident they are all referring to the same thing.
>
> Ensembl goes to great lengths to try to maintain stable names for genes and other features in the genome."[8]
>
> (Ensembl Naming Conventions. From Ensembl website database.)

The diversity of kinds of entity and the desirability of taxonomic stability seem exactly to mirror the issues that arise for the case of species. So my proposal is for an atheoretical pluralism similar to that which I advocate for species: a gene is any bit of DNA that anyone has reason to name and keep track of. Genes may be proper parts of other genes; they may overlap; they may have non-contiguous parts, perhaps on two or more chromosomes. And, as illustrated for the case of developmental defect genes, "gene" may even refer to a functionally connected class of DNA segments. My conclusion is that there are genes—an important point given how much people talk about them—but that the price of this is conceding that it doesn't take much to be a gene. Not much, but not nothing either. I am assuming that genes are real material entities. Many of the genes discussed by behavioral geneticists for instance, may well not even meet this minimal condition.

Some Consequences

One conclusion I would like to draw from this recognition of the diversity of the referent of the term "gene" is a familiar one in contemporary philosophy of biology. It is that the traditional philosophy of science that sees science as ultimately concerned with the articulation of wide-ranging and fundamental truths—laws of nature, for instance—has little relevance to biology. Genetics and genomics offer little project of such general truths because of the diversity of their subject matter. Recalling Rheinberger's suggestion mentioned above, it may be that the function of the most general terms in such sciences is precisely to compensate for the lack of such general truths by allowing some degree of communication between people

8. http://www.ensembl.org/. Thanks to Dick Holdsworth for drawing my attention to this statement.

with varying interests in the workings of, for example, the genome. Similarly, I suggest, the term "species" is useful in allowing people with different interests in the classification of organisms, different principles for accomplishing this, and consequently different groups of organisms to which they need to refer, nevertheless to understand when reference is being made to a group of organisms at a certain important level in the taxonomic hierarchy. The great diversity of the subject matter of biology calls for the most central terms not to be those in terms of which laws can be formulated, but rather those which are tolerant enough in their reference to bridge the divides between the various phenomena in which local communities of researchers may be interested. There are, I suppose, some general truths about DNA that make it possible for DNA to constitute genes, but there are lots of ways for bits of DNA to be genes of various kinds, and all of these depend on the relations between bits of DNA and other things to which they are related.

It is, as I have noted, hardly a novel suggestion that the view of science as the search for universal laws is of little or no relevance to biology, but the extent to which this suggestion has been reinforced by recent developments in genetics has not yet been fully appreciated. Indeed, it is still sometimes imagined that the annoying failure of biology to generate law-like generalizations is a consequence merely of its continuing concern with complex and variable structures, and its concomitant failure to get down to the real action at the molecular, and ultimately even more fundamental, levels.

One moral of my preceding remarks is just that no such consequences result as we investigate the inner structures of biological things. On the contrary, what we find as we become more familiar with molecular processes is a diversity of structure and action quite comparable with that which we find at more complex levels. We are far from approaching the few simple laws that earlier theorists imagined might reduce complexity and diversity to order and uniformity.

My argument here is not in any simple way anti-reductionist. It is clear in genetics that enormous illumination and insight has come from our ability to investigate and describe molecular processes. It is, however, anti-reductionist in the sense of rejecting the hierarchical view of nature often associated with reductionism. Knowledge of different levels of organization is complementary, not competing. The molecular view is not a superior view to, say, the cellular view, and one that in principle should render the latter obsolete. And the reason for this is simply that the molecular view is not even separable from the cellular view. There is no possibility of specifying the behavior or function of bits of DNA independently

of a detailed description of the biological context in which they exist. Minimally this context will include further genomic and cytological information. Sometimes the relevant context will be much broader, including physiology, ecology, and even sociology. And of course this dependence on context is a large part of why what may look very similar—strings of DNA—may nevertheless prove to be so diverse.

There is a vision of the cell as a nugget of information suspended in a soup of dumb and formless goo, a notion that still seems common in popular presentations of biology, and this vision perhaps best represents the remaining aspirations of hierarchical reductionism. The extranuclear goo, in this vision, is no more than the minimal context necessary for the expression of the structure inherent in the DNA.

But in reality the extranuclear goo is as structured, as rich in information, as is the nuclear DNA. The sorts of things bits of DNA can do involve diverse reactions with particular chemicals and structures in the cell. Biochemistry only becomes molecular biology when it is embedded in cytology. Lower level knowledge cannot possibly displace higher level knowledge.

And this, as one final important philosophical moral of our growing understanding of the cell and the genome, should also make clear the futility of seeing causality as something always elusively located in lower levels of order, ultimately filtering up from the most basic constituents of matter. What was once the controversial thesis of "downward" causation is a commonplace in biology. One striking example is the differentiation of cells in development. All the diverse varieties of cells in multi-cellular organisms, the liver cells, blood cells, hair cells, and so on, trace their origin to the same ancestral cell. The explanation of the different developmental paths leading to these diverse outcomes does not reside in differences in the DNA, but in the ability of the spatial relations between cells and the spatial distribution of relevant biologically active substances in the egg and, later, in different locations in the body to affect differentially the behavior of the DNA within different cells. This seems as clear as possible a case of the behavior of a low level entity being caused by higher level entities of which it is part. The prejudice in favor of the causal priority of the small, visible in a range of weak reductionisms and supervenience theses is, I think, just that, a prejudice.

I don't know how useful it is to read scientific models as political allegories. But it is remarkable how naturally a common picture of the cell fits with a hierarchical model of social organization. Command and control inheres in a central administration, the genome, and orders are carried out

by messengers, clerks (transcribing, translating, and so on). The construction work takes place at various sites decently removed from the seat of power. Contemporary molecular genetics takes us away from this Stalinist model towards something more Smithian. The efficiency of the cell is unimaginable, from this perspective, without the distribution of specialized capacities across a very large range of different agents. Command and control do not descend from the central administration building, but emerge spontaneously, as if guided by an invisible hand. On the Smithian model order at the lower level is an order of teleological mechanism: events fit together in efficient ways to produce valued outcomes. Broad generalizations—like this last one—emerge if at all at higher levels. Empirical evaluation of the attempt to provide a science of economics as a set of axioms and their consequent theorems shouldn't encourage us to hold out a great deal of hope for these higher level generalities.

Genetic things, genes in the catholic sense I have advocated, are unquestionably real. They cause things to happen at the phenotypic level and intervening levels of biological structure, just as those things cause the activation and specific action of particular genes. Hence only at many levels simultaneously can we begin to get a full account of the nature of an organism.

One final question, the answer to which I hope has been illuminated to some degree by the foregoing, is why so much contemporary discourse is replacing the term "genetics" with "genomics." Genetics, a science of hypothetical entities held to be responsible for inheritance, can be caricatured, but not altogether unfairly, as a science developed in accordance with a reductionist epistemology and a law-seeking methodology. Over the course of a century genetics led us to a remarkably detailed view of the genome. Among many remarkable properties of the genome is its total unsuitability for both this epistemology and this methodology. Genomics, I am tempted to suggest, is the successor science to genetics that rejects this obsolete epistemology and methodology.

A more positive way of stating the point recapitulates and largely endorses Rheinberger's idea. Traditional philosophy of science sees central concepts as opening up the possibility of discovering laws of nature or, at any rate, general knowledge of nature. But as I have tried to show, genomic events are diverse and specific and there are few if any laws that apply to genes. The example of genes suggests something quite different: the function of this concept is rather to allow us to talk about lots of different things. Such a concept facilitates communication between people with different but related concerns, and facilitates continuity between successive historical inquiries. It also provides a risk of serious misunder-

standing. This risk is probably minimal in the case of working scientists communicating their results to one another, as far more specific, local interpretations of a word such as "gene" will be expected and provided. Misunderstanding arises rather as scientific results disseminate either to different areas of science or to other domains of human life, and such dangers may be exacerbated by obsolete philosophy of science. At any rate, there is a likely role for philosophers of science in attempting to delineate the diversity of meanings of such complex concepts.

References

Beurton, Peter J. 2000. "A Unified View of the Gene or How to Overcome Reduction." Pp. 286–314 in *The Concept of the Gene in Development and Evolution.* Edited by Peter Beurton, Raphael Falk, and Hans-Jorg Rheinberger. Cambridge: Cambridge University Press.

Dupré, John. 2002. *Humans and Other Animals.* Oxford: Oxford University Press.

Fogle, Thomas. 2000. "The Dissolution of Protein Coding Genes in Molecular Biology." Pp. 3–25 in *The Concept of the Gene in Development and Evolution.* Edited by Peter Beurton, Raphael Falk, and Hans-Jorg Rheinberger. Cambridge: Cambridge University Press.

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

Ikemura, Toshimichi. 1981. "Correlation between the Abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system." *Journal of Molecular Biology,* 151: 389–409.

Johannsen, Wilhelm. 1909. *Elemente der exakten Erblichkeitslehre.* Jena: Gustav Fischer.

Lakatos, Imre. 1980. *The Methodology of Scientific Research Programmes. Philosophical Papers: Volume 1.* Edited by John Worrall and Gregory Currie. Cambridge: Cambridge University Press.

Lloyd, Elisabeth. 1994. *The Structure and Confirmation of Evolutionary Theory.* Princeton: Princeton University Press.

Morgan, Thomas H., Alfred Sturtevant, Hermann J. Muller, and Calvin Bridges. 1915. *The Mechanism of Mendelian Heredity.* New York: Holt.

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

Rheinberger, Hans-Jorg. "Gene Concepts: Fragments from the Perspective of Molecular Biology." Pp. 219–39 in *The Concept of the Gene in De-*

velopment and Evolution. Edited by Peter Beurton, Raphael Falk, and Hans-Jorg Rheinberger. Cambridge: Cambridge University Press.
Waters, Kenneth. 1994. "Genes made molecular." *Philosophy of Science,* 61: 163–185.