This chapter to be published as:

Sapienza, Carmen (2009). “Selection Does Operate Primarily on Genes: In Defense of the Gene as the Unit of Selection” in Ayala, Francisco and Arp, Robert, *Contemporary Debates in Philosophy of Biology*. Oxford: Wiley-Blackwell.

Selection Does Operate Primarily on Genes: In Defense of the Gene as the Unit of Selection

Carmen Sapienza

Natural selection is an important force that shapes the evolution of all living things by determining which individuals contribute the most descendents to future generations. The biological unit upon which selection acts has been the subject of serious debate, with reasonable arguments made on behalf of populations, individuals, individual phenotypic characters and, finally, individual genes themselves. In this essay, I argue that the usual unit of selection is the gene. There are powerful logical arguments in favor of this conclusion, as well as many real-world examples. I also explore the possibility that epigenetic differences between individuals may be heritable between generations. Although few such examples exist, epigenetic differences provide an exciting source of potentially heritable variation that may allow rapid evolutionary change to occur, perhaps in response to environmental influences.

1. Introduction

Natural selection may be defined as a mechanism that distinguishes differences between biological entities and results in a net reproductive advantage for one of them. I will assume that natural selection is a significant force in evolution and will not debate whether it is the only force shaping evolutionary change. I will defend the idea that the “gene” is the usual and most important level at which natural selection distinguishes differences between biological entities. The notion that the gene is the unit of selection assumes, further, that differences between

genes underlie almost all forms of heritable variation. In this chapter, I have attempted to argue from first principles, accompanied by a few real-world examples that I believe make the case that natural selection has shaped important complex traits and that these traits are controlled by one or a few genes. This form of argument has been aided considerably by Professor Burian's thorough and thoughtful discussion of the history of this debate in the companion chapter. I refer the reader to Professor Burian's essay and wish the reader to know that I am in substantive agreement with much of his discussion.

2. Natural Selection Operates *within* Genomes without Regard for Phenotypic Effect

I suspect that the reason I was invited to defend the idea that natural selection operates primarily on genes is because W. Ford Doolittle and I co-authored a highly-controversial article in 1980, entitled "Selfish genes, the phenotype paradigm and genome evolution" (Sapienza & Doolittle, 1980). In that piece – and a companion article written by Leslie Orgel and Francis Crick (1980) – we extended Richard Dawkins' (1976) selfish gene argument to the level of genome structure. We argued (correctly, I believe, to this day) that much of present day genome structure is the result of natural selection operating directly on DNA sequences for the capacity to make more than one copy of themselves prior to cell division/meiosis. In other words, many elements in the genome are present simply because they have the capability of making copies of themselves and

spreading these copies around the genome. In conjunction with sexual reproduction, such behavior becomes the equivalent of "meiotic drive" and the new copies of the elements will spread throughout the genome, much like an intra-genomic parasite. In fact, members of a small number of families of these elements make up more than a third of the human genome and some individual families have more than a million members. Such sequences are sometimes erroneously referred to as *junk DNA*, inferring that the sequences have no function (as far as individual phenotype is concerned). A less anthropomorphic explanation is that their function is to make more copies of themselves (much like the function of a virus is to infect cells and make more viruses). In rare cases, insertion of an element into or near a gene may have an effect on the phenotype of the individual (and so be subject to natural selection operating on organismal phenotype); however, it is impossible to imagine that such has been the case for all one million accumulated members of the human Alu family (Batzer & Deininger, 2002), for example, in so far as most of them are not present within or adjacent to genes (Ensembl, 2007). I have always found this logical argument compelling, in the extreme, and have concluded that the vast majority of these "transposable elements" have survived and increased in number within genomes largely in the absence of supervision by any selective force operating on organismal phenotype.

In the remainder of this chapter, I will argue that there is at least one additional way that natural selection can, and does, operate directly on genes – via non-random segregation of chromosomes – and that, even in those cases where natural selection appears to operate on some complex phenotypic difference between individuals, the difference is most likely traceable to genetic differences at one or a few loci.

3. Selective Forces, Heritable Variation, and the Definition of Function

During my graduate student days, I was fortunate to attend a small meeting on genome evolution at which the late John Maynard Smith was a featured speaker. Professor Maynard Smith was a wonderfully eloquent communicator who had the ability to reduce complex problems to manageable components. His definition of the term *function* has served as a guiding principle in my attempts to explain biological variation throughout my career. He noted that when an evolutionary biologist made the statement "the function of the heart is to pump blood," what he/she actually meant was *not* simply that the heart *did* pump blood but that those individuals whose hearts were superior in pumping blood survived and left more descendants than those individuals whose hearts were inferior in this function (for more on this, see the chapters by Perlman and Cummins & Roth in this volume).

As was the case with many of Maynard Smith's simple examples, a complex web of cause and effect was concealed just below the surface. There is the assumption that hearts of different blood-pumping abilities are carried by

different individuals in the population and, further, that the different blood-pumping abilities of the different hearts is heritable, so that individuals with superior hearts are more likely to have offspring with similarly superior hearts than are individuals who have hearts of inferior blood-pumping ability. Layered on top of these caveats is the question of how, exactly, individuals with superior hearts come to leave more offspring than individuals with inferior hearts. Is it because they can run faster or for longer distances, thus escaping predators? Or is it because they are less likely to die as a result of myocardial infarction and so have a longer reproductive lifespan? There are many possibilities but the gist of determining the "function" of biological structures or processes is the formal identification of selective forces and the determination of how each force distinguishes between variants.

So, in Maynard Smith's example, we would be left with the question of what the selective force "sees" and whether what is seen is attributable to one gene, a few genes, or some higher collective property of the organism. Of course, it is true that a complex organ, like a heart, is not seen by selection on its own but in the context of the creature bearing that heart. Having a heart attack while running the Boston Marathon cannot be ascribed to variation in a single gene – or can it?

My argument is that the things about heart performance that are likely to matter most – for example, systolic and diastolic blood pressure, likelihood of

myocardial infarction and serum cholesterol levels – all show very high heritability (Jorde, Carey, Bamshad, & White, 2000). In other words, the variation that is seen in blood pressure between different individuals in the population can be explained, in large part, by differences in genotype. The number of genetic differences required to explain these phenotypic differences is not known precisely. However, millions of additional years in patient life-span (not to mention, billions of dollars in drug company profit) have been realized as a result of treatment of two of the most common and dangerous cardiovascular phenotypes: hypertension and high serum cholesterol. Both of these conditions can be alleviated by drugs (ACE inhibitors and statins) that target the products of single genes. Insofar as variability in cardiovascular phenotype is a product of heritable variation (even those of us who are loathe to be gym rats must admit to some environmental effects), it seems probable that the phenotypes under selection are controlled by small numbers of genes.

4. Natural Selection Can, and Does, Act on the Products of Individual Genes

Given the complexity of living organisms and the likelihood that many phenotypic characters are the result of the action of multiple genes, it is worth entertaining the question of whether there are examples of natural selection acting at the level of individual genes. Simple, real-world examples from bacterial genetics come to mind; for example, the colonization of hospitals by microorganisms that are resistant to various antibiotics is a clear case where

organisms that differ only by the acquisition of a single gene come to dominate an environment in which both might persist for many generations in the absence of the selective agent. Because the enzymes that break down penicillin, for example, tend to be shuttled from bacterium to bacterium on easily mobilized genetic elements – small pieces of DNA, called *plasmids* and *transposable elements*, that contain one or a few genes – the accumulation of antibiotic-resistant microorganisms in hospitals has occurred with breathtaking speed.

In such cases, the selective force (the presence of antibiotics) has distinguished between, and among, genetic variants by killing those that do not produce the product of the gene encoding antibiotic resistance. In fact, in this instance what is being selected is not the gene, *per se*, but the presence of the product of that gene. While there is only minimal argument that without the gene, there would be no gene product, the reciprocal statement is not true; there are many examples in which a particular gene is present but the gene product is not. The most obvious examples are carried around with each of us while we go about our daily business. Even though each of our 10^{14} cells contains all ~25,000 of our genes, every cell is not producing all of the gene products of all of our genes. The processes of development and differentiation lead to these characteristic epigenetic differences between cells that are not based on genotypic differences (with a few notable exceptions). Instead, the DNA of liver cells is packaged differently within the cell nucleus from the DNA of brain cells and the selection

of genes that is accessible to the RNA transcription machinery is different in each type of cell. Given that it is possible for genes to be present in cells (including sperm cells and egg cells) without expressing a gene product, it is worth asking whether it is possible for natural selection to operate on genes, themselves, directly.

5. Natural Selection Can Act Directly on Genes Themselves

Perhaps the most convincing demonstration that natural selection may operate at the level of individual genes comes from examples of *true meiotic drive*, i.e., non-random segregation of a chromosome at meiosis. The process by which an egg is formed in females of most species is an asymmetric meiosis. Instead of producing four gametes that are equal in size and genetic content, females produce eggs by producing one large gamete (the egg) and two smaller meiotic products (the 1^{st} and 2^{nd} polar bodies). The genes contained in the egg have the possibility of contributing to the next generation, while the genes in the polar bodies do not.

We (Wu et al, 2005), and others (Agulnik, Agulnik, & Ruvinsky, 1990), have found that it is possible for individual chromosomal variants to influence the probability that they are segregated to the egg-half of a meiotic division directly – and, thus, their survival for another generation – rather than the polar body half, where they have no chance of being represented in the next generation. In most of these instances, the chromosomal structure upon which natural selection is

operating is the *centromere*, i.e., the complex of DNA sequences and proteins responsible for attaching the chromosome to the meiotic spindle (the cellular structure responsible for the actual partitioning of chromosomes between daughter cells) so that the spindle microtubules can move the chromosome to one pole or the other.

When chromosomes are observed under the microscope, they appear as strands of nearly uniform thickness, except for a single constriction called the *centromere*. The centromere is the structure responsible for the physical movement of the chromosomes between daughter cells at cell division. The centromere contains proteins to which microtubules (cellular "motors" responsible for the movement of many cellular components and proteins) attach and enable the chromosomes to be moved away from the plane of cell division and ensuring the orderly and equal segregation of the chromosomes to each daughter cell.

Evidently, all chromosomes have centromeres, but not all centromeres are equal in their ability to attach a chromosome to the egg-side of the spindle (Pardo-Manuel de Villena & Sapienza, 2001a). Interestingly, such *meiotic drive* occurs even when the outcome is disadvantageous to the organism. In the case of the *Ovum mutant* allele carried by the DDK inbred mouse strain, for example, inclusion of the DDK allele in the egg risks embryonic death if the egg is fertilized by sperm from some other inbred strain males; yet the DDK allele is

included in the egg more than 60% of the time in females who are heterozygous for a DDK allele and a "wild-type" allele (Pardo-Manuel de Villena, de La Casa-Esperon, Briscoe, & Sapienza, 2000; Wu et al., 2005). Even in the face of a net reduction in the overall fitness of individual heterozygotes, natural selection (in the form of the DDK allele segregating, preferentially, to the egg) is predicted to result in a net increase in the fraction of DDK alleles in the population, simply because a higher fraction of eggs will contain DDK alleles.

Before our laboratory started working on this problem, I had been under the impression that examples of meiotic drive were rare. I was wrong. One of the more interesting examples of biological variation, without direct phenotypic consequence to the carrier, is the formation of *Robertsonian translocation chromosomes*. Such chromosomes are end-to-end fusions of smaller *acrocentric* chromosomes (each having a centromere at the tip of the chromosome) into a larger chromosome with two arms and a centromere in the middle. These chromosomal variants are created with extraordinary frequency: approximately 1 per 1,000 meioses in human females (which is several orders of magnitude greater than the frequency of any other genetic event) and result in a meiotic pairing configuration in which one member of the pair must attach a single centromere to the meiotic spindle, while the other must attach two centromeres to the spindle (see Hamerton, Canning, Ray, & Smith, 1975). As a group, mammals have an unusually constant genome size of approximately 3×10^9 base pairs per haploid

genome but vary widely in the number of chromosomes over which the genome is distributed, from 6 chromosomes (in the Indian muntjac deer) to 102 chromosomes (in a small South American rodent) (Scherthan, 2007). As it turns out, cycles of chromosome fusion/chromosome breakage and meiotic drive have shaped all of mammalian karyotypic evolution (Pardo-Manuel de Villena & Sapienza, 2001b), with individual karyotypes being driven to favor all, or most, chromosomes to be of one form (i.e., acrocentric, as in many mice) or the other (metacentric, as in humans and great apes).

This variation in chromosome form has proven a major mechanism of reproductive isolation (and, therefore, speciation) in mammalian evolution. Individuals from different populations that carry/do not carry the fusion chromosome are often perfectly interfertile. However, the resulting heterozygotes often have meiotic segregation difficulties, resulting in aneuploid gametes (having an abnormal number of chromosomes) with phenotypic consequences for the next generation (Gropp & Winking, 1981). In fact, closely related species exhibit such karyotypic differences far more often than might be expected by chance. A notable example is that human chromosome 2 appears as two separate chromosomes among the karyotypes of the great apes (Dutrillaux, 1980).

These examples of direct selection at the level of the gene (as represented by genetic variability in the structure of centromeres) affect reproductive fitness, which is the most important outcome measure in any discussion of population

genetics. Although one might hedge that centromeres, themselves, are epigenetic structures, the innate biological variability between centromeres that allows them to be subject to meiotic drive is fundamentally genetic (Henikoff & Malik, 2002).

6. What Are the Limitations on the Unit of Selection Being "the Gene"?

I have a wonderfully clever and inventive colleague who works on G-protein-mediated signaling. We are often like-minded, politically and socially, and agree on many divisive subjects that might bring others to blows. However, we have never seen eye-to-eye on the "function" of G-proteins. Fortunately, we are each dismissive of the others' views on this topic, rather than confrontational, and so our disagreement does not cause much friction between us. Our disagreement stems, I think, from our views on natural selection and how natural selection might affect the function of G-proteins. My colleague believes that it is likely that selection has optimized all of the G-proteins to serve a unique function. I do not.[1]

This difference of opinion is likely to reflect our level of biological focus. My colleague would characterize himself as a molecular biologist. I characterize myself as a geneticist. Both of us are comfortable with the notion that natural

[1] G proteins mediate cell signaling via guanine nucleotide binding and additional couplings with more than 800 different receptors (in humans). Cell signaling is a crucial biochemical process by which signaling molecules bind to receptor molecules that are specific and the binding of the signal to the receptor keys a change in a biochemical pathway. The details of the biochemistry are not important for this discussion (although they were important enough that their discoverers were awarded a Nobel Prize in 1994).

selection can act on genes or, in this case, gene products. However, my colleague is much more reductionist than I am, and sees *all* genes or gene products as the product of natural selection. I am of the opinion that traits controlled by large numbers of genes are difficult, if not impossible, for natural selection to "see." I believe this must be so because, although it is likely that selection can optimize a particular subunit of a G-protein to bind GTP, I think that the very large number of combinations of protein subunits observed to participate in various forms of G-protein-mediated signaling is too large for natural selection to "see."

My colleague tells me that functional G-proteins are *heterotrimeric*, meaning that the functional proteins contain alpha, beta, and gamma subunits and that G proteins are encoded by a family of thirty-five genes: sixteen alpha, five beta, and fourteen gamma, all of which are scattered over several chromosomes (Dutrillaux, 1980). The interplay between various G-proteins and various receptors gives rise to a large number of complex phenotypic traits, many of which can be seen, with little imagination, to have adaptive significance. Our ability to discriminate between the odor of roses and the odor of manure, for example, is the result of the interaction of particular G-proteins with particular olfactory receptors in particular olfactory neurons. One might think, then, that particular G-proteins would have evolved to play specific roles.

However, my argument, simply stated, is that even if rare variants of a particular alpha, beta and gamma subunit were to come together in an individual

to reward him/her with the most exquisite sense of smell (rather like the character in the popular novel, *Perfume* (Suskind, 1986), the allelic combinations giving rise to this trait would be broken in the next generation because the genes are on different chromosomes and segregate independently. If only three unlinked genes were involved in this trait (a particular alpha, a particular beta and a particular gamma subunit, for example), then ½ x ½ x ½ = 1/8 of gametes would carry the proper combination. Going beyond three genes requires ever greater population sizes and numbers of offspring or ever greater selective advantage for the variant trait, in order to bring natural selection face-to-face with the trait. If we use the minimum reproductive advantage necessary to maintain an advantageous trait in the population (we may use the one given in Professor Burian's chapter: 1 additional offspring in ~10^4), then natural selection cannot, by definition, "see" beyond ~13 independently segregating genes (2^{13} = 1 in 8096).

In practice, the minimum reproductive advantage required to maintain the trait is likely to be substantially greater, given changing environments or weak selection, reducing further the number of gene variant combinations that natural selection can see. The easiest way to get around this problem is to create situations in which the desirable gene variants do not segregate independently: to "link" the gene variants affecting the desirable trait on the same chromosome in gene clusters. Numerous examples of this strategy are available, including the large family of olfactory receptor genes whose signals are transduced by the G-

protein signaling molecules with which we opened our discussion. If an advantageous combination of variants alleles at two separate genes were to become closely linked (as a result of a chromosome rearrangement, for example), they will ensure the inclusion of the advantageous combination in ½ of gametes rather than ¼ of gametes (if unlinked). Given the demonstrated tendency for genomes to link genes that work in the same developmental or biochemical pathway in this way, I would argue that this outcome is not required if the unit of selection is something other than the gene but is a prediction of the hypothesis that the unit of selection *is* the gene.

7. The "Complexity" Argument: Do Complex Phenotypes Require Complex Explanations?

The notion that natural selection "sees" only traits that are the result of the action of multiple genes (i.e., organs or organisms) seems, to me, a bit like the irreducible complexity argument used by supporters of creation science. The gist of their argument is as follows: the eye is a very complex structure that is able to receive visual stimuli and transmit those signals to the brain where they can be processed into information upon which decisions that affect fitness may be made (climb the tree and avoid the wolves, stay on the ground and fight, or run?). If any one part of the eye were removed/did not function properly, it would fail to fulfill its function and, thus, the eye could not have been designed by natural selection but must have been designed by an intelligent designer. In the same

way that I would argue that a badly functioning eye is better than no eye (ask any visually-impaired person whether they would prefer to be completely blind or badly impaired), I *do* see how it is possible to add/subtract layers of functional complexity by changing one gene at a time.

I think that the ultimate argument under this heading is illustrated by the ultimate quantity that all evolutionary arguments must take into account when hypothesizing a selective advantage/disadvantage for *any* trait: biological fitness. At bottom line, fitness is simply the number of offspring provided to the next generation. Fitness is the sum-total of all of the biological, social and environmental variables at operation during the life span of any organism. One might argue that fitness in the human population would be an incredibly complex issue, affected strongly by economic, social, environmental and biological variables. I would have predicted, *a priori*, that tracing reproductive success to any particular variable would be impossible. Fortunately, my certainty on this subject can be listed under the comment of my former supervisor that "not everything I know is true."

A few years ago, a company called Decode Genetics was formed in Iceland, with the intention of "mining" the genetic variability of the Icelandic population in order to find genetic variation associated with common diseases. The social and political structure of Iceland makes such an endeavor easier, in some ways, than it might be in a more diverse population. The integration of

birth and medical records with genotype information has made it possible to analyze fitness among virtually the entire population of Iceland as well as ask whether there are genetic factors that are correlated with fitness (Kong et al., 2004). The conclusion, from analyzing more than 14,000 offspring in more than 5,000 families, is that the women who had the largest numbers of offspring (women being the important variable) were those who had managed to reproduce at the oldest ages. In other words, all things being factored out across the entire Icelandic population, the women with the longest reproductive lifespan had the most children.

This may not seem terribly surprising, upon short reflection. However, what *was* surprising, to my mind, is that fitness correlated with the number of recombination events observed in their offspring. Females who had higher levels of recombination were able to reproduce at older ages, and had more offspring, than females who had lower levels of recombination. Because failure of recombination is a risk factor for aneuploidy – and 50% of spontaneous abortions are aneuploid (Hassold & Hunt, 2001), so this factor has a major effect on reproductive success – the suspicion is that the ova of females with more recombination events are less likely to be aneuploid at older ages than the ova of females with fewer recombination events.

It is possible that many genes could affect recombination rate, but it is certainly true that we, and others, have demonstrated inter-strain differences in

recombination that are attributed to single loci (specifically, in mice) and that there are many cases of one or a few genes affecting recombination rates dramatically in many organisms (de La Casa-Esperón et al., 2002; Kong et al., 2008). Overall, I would argue that, while it is certainly possible to envision situations in which the unit of selection might be something much more complex than an individual gene or small numbers of genes, the availability of real-world examples of the opposite tendency make me question the wide-spread utility of more complex explanations.

8. Do "Epigenes/Epialleles" Provide a "Non-genetic" Source of Heritable Variation Upon Which Natural Selection May Act?

Because I have taken the position that all or most heritable biological variation is based on true genetic differences between genes, it is important to address the possibility that variation caused by epigenetic differences is also heritable. By my definition, very little of this form of biological variation is heritable, because by *heritable*, I mean the variant epigenetic form must be transmitted to the next generation, unaccompanied by a causal genetic difference.

Most epigenetic differences survive somatic cell division. Somatic cells are all of the cells of the body except for sperm cells or egg cells. Indeed, faithful replication of epigenetic differences is the basis for differentiation and development; for example, progenitors of liver cells continue to produce liver cells and not brain cells (and it is this form of programming that is at the practical

root of much of the embryonic *versus* adult stem cell debate). The phenotypic difference between these genetically identical cells is based on the somatic heritability of *chromatin* structure. Chromatin is the complex of DNA, histone proteins, and other proteins, that bind to DNA and small molecular modifications of DNA (the addition of methyl groups to certain combinations of letters in the DNA code) and modifications of histone proteins (the addition of methyl or acetyl groups to certain amino acids). The DNA of liver cells is packaged into chromatin differently than the DNA of brain cells, enabling different groups of genes to be expressed in each cell type. Liver cell-specific chromatin packaging is replicated from one liver cell division to the next. However few, if any, of these epigenetic differences are transmitted through the germline to the next generation. There are a few examples of such "transgenerational" epigenetic inheritance (Rakyan et al., 2003) but most are also accompanied by genetic differences (Chong, Youngson, & Whitelaw, 2007).

Nonetheless, I must admit to being intrigued by the formal possibility of transgenerational epigenetic inheritance. It is certainly possible, if so far rare, for epigenetic variability to provide a *non-genetic* form of biological variability. This form of variability could provide a mechanism for the inheritance of acquired characteristics: if a particular environmental factor resulted in selection for turning on a particular combination of genes early in development and these genes were newly expressed in both brain and testes, for example, it is also possible that

particular behavioral patterns dictated by the newly expressed genes could be programmed in the next generation by their chromatin packaging in some fraction of sperm cells. The demonstration of such a development would be exciting, indeed.

9. Summary Points: The Usual Unit of Selection Is the Gene

The following is a summary of the main points of this chapter: (1) much of complex genome structure reflects the accumulation of mobile elements that increase in number, regardless of effect on organismal phenotype; (2) there are examples of selection leading to replacement of one population of organisms by another – viz., near isogenic population – that differs from the original at only a single or small number of loci; (3) selection, *via* meiotic drive/non-random segregation of chromosomes, can act directly on DNA sequences or chromosomal structures in the absence of gene products encoded by those particular sequences; (4) a substantial fraction of the observed variation in complex traits that are likely targets of selection can be traced to genetic differences at a small number of loci; (5) even under the most generous estimates for how small a difference between entities natural selection can "see," no trait that is the product of more than 10-12 unlinked genes can be selected; (6) the problem of independent segregation of unlinked genes predicts that if traits are encoded by multiple genes, those genes will tend to become associated in fewer linkage groups and become easier for selection to see; (7) although epigenetic variation would provide a rich source of

potentially heritable, but non-genic variation, very few examples of transgenerational epigenetic inheritance have been documented. Most such differences are accompanied by underlying genetic changes or are strain/population specific.

Postscript: Counterpoint

Of the arguments raised by Professor Burian against the idea that the usual target of natural selection is the gene, there are a small number with which I disagree, on the basis of the evidence or strength of argument. On the other hand, I am in agreement with Professor Burian that there is one issue that does require (or, at least, may require) selection to act on something other than individual genes. First, let us look at our disagreements, as these are the easiest to address.

1. What *Does* selection at the Level of the Gene Explain?

Professor Burian takes issue with the notion that quantitative traits – such as height, blood pressure, and serum cholesterol – that are influenced by environmental factors as well as genetic factors, could trace the genetic component of their variance to a small number of genes. He is correct in that, indeed, there may be hundreds of genes that influence these phenotypes. However, my point is that the bulk of the variance, in these and many other quantitative traits, is attributable to variation in a small number of genes. The simple proof of this statement is that while diet and exercise undoubtedly influence hypertension and/or hypercholesterolemia, a large fraction of the

hypertensive or hypercholesterolemic population has altered their blood pressure or their serum cholesterol by taking drugs that target the product of single genes. Does this mean that there is only one gene that controls blood pressure or serum cholesterol? The short answer is no, but there are individual genes that have a disproportionate effect. Most models of quantitative trait variation suggest that multiple genes are at work and that their effects are *additive*, i.e., genetic variation at genes "1" through "n" results in the total genetic variance observed. However, the effect of each gene need not be equal. Although genetic variation at twenty loci may contribute to the total genetic variation, fifty percent of the trait variance may be due to the effects of two genes, with the remaining eighteen genes each contributing a small amount to the remaining fifty percent of genetic variation. Under this scenario, selection may act strongly on the two genes with large effect and weakly, or not at all, on the remaining eighteen.

Additional evidence that multiple genes with unequal effect is a common feature of quantitative traits comes from whole genome association studies of other phenotypes. The completion of the human genome sequence and resequencing efforts to determine the amount of genetic variability between individuals has allowed geneticists to take an unbiased approach to determining what fraction of phenotypic variation maps to genetic variation in individual genes. Height, as Professor Burian notes, is a good example of a trait controlled by the effects of many genes. However, with respect to individual populations

whole genome association analysis has found that, although height is a polygenic trait, there are likely a few genes that explain the majority of the variation, while many other genes contribute to background "noise" (Liu et al., 2006). Different genes may be found in different populations (Perola et al., 2007), but whole genome association analysis, by nature, would not be so successful in finding genes with significant effects on these phenotypes if they were truly the result of hundreds of genes, each with equal effect.

In fact, I believe it is this difference in focus that underlies our disagreement over the usual target of selection. If there are complex regulatory networks underlying most phenotypes, then my argument that natural selection cannot "see" so many genes (if they are unlinked) can be used as an argument that selection does not operate on genes. If, on the other hand, most phenotypic differences in complex traits have smaller numbers of genes at their root, there is no need to postulate that selection operates at a higher level. My view on this difference is reminiscent of the complex charts detailing the reactants, enzyme catalysts and products of the linked biochemical reactions that make up glycolysis or the Calvin cycle or Kreb's cycle. Anyone who was forced to memorize the intricacies of these charts as a student also remembers that there were "rate-limiting" steps in most of these pathways. While all steps were important, the flow of reactants and products through the pathway was not controlled equally at every step and mutations in genes in some steps in the pathway had a much

greater phenotypic effect than mutations in others. A less academic example is the towing of cars stopped on the shoulder of the highway. While automobiles are enormously complex and have thousands of moving parts, the selection of any particular car by a tow truck does not often have to do with the failure of many parts (barring a crash) but is usually the result of failure of a particular part in one system or another and each system (ignition, transmission, steering, braking, etc.) tends to have a hierarchy of components that are most frequent to fail.

My last minor disagreement with Professor Burian is over the role of epigenetics in producing heritable variation. I believe it is possible, and would be quite exciting, if epigenetic variation was found to provide an important "non-genetic" source of heritable variation. In the case of the particular trait that I addressed earlier – viz., the contribution of genetic *versus* epigenetic variation to differences in recombination rate – the difference between the male recombination rate and the female recombination rate may be almost entirely epigenetic. However, it is most likely that the inter-individual variation between females (which is the variable of interest in this case) is due to true genetic variation at loci involved in DNA repair rather than to epigenetic differences between females.

2. What *Doesn't* Selection at the Level of the Gene Explain?

Professor Burian makes a valid point that there are cases in which it is difficult to argue that selection is operating on a particular gene or allele because the

phenotype under selection is elicited only in combination with another gene or allele. Examples of *hybrid vigor* or *overdominance* (like Professor Burian's sickle cell example) have long fascinated geneticists because they represent cases in which a trait that is present in neither parent shows up in the offspring. However, it is worth pointing out that hybrid vigor (and cases in which the hybrid is less fit than either parent) is a relatively rare circumstance. Most hybrids, in fact, have phenotypes that are similar to one parent or the other (dominant traits) or some average/intermediate value between the two (additive traits).

Because the goal of this exercise was to present arguments for what is the usual and most likely target of selection, I believe some quantitative data are in order. The sequencing of human, mouse, and other animal genomes, has been accompanied by the development of genome-wide and high-throughput methods in which it is possible to determine which alleles of which genes are being transcribed in tissues or individuals. Such transcription profiles may be compared between any two parents and their offspring to determine whether the amount of transcript of any particular gene resembles that of one parent or the other or is some average between the two. Although the quantitative trait being analyzed in this case (amount of transcript produced from an individual allele) may not always be the best measure of a phenotype produced by any gene, it is a useful phenotype for which one may ask, of thousands of individual genes, whether it is heritable, dominant, additive, or overdominant.

The most comprehensive analysis provided, so far, is from the work of Cui, Affourtit, Shockley, Woo, & Churchill (2006). These investigators compared transcript levels in two inbred strains of mice and their reciprocal F_1 hybrids. Nearly 9,000 transcripts showed evidence of heritability in transcript level, with a median heritability of ~70%. Approximately, 20% of the heritable transcripts exhibited dominance (levels similar to one parent), while the bulk of the remainder showed additive inheritance. Only 167 transcripts (less than 2% of the total heritable transcripts) exhibited overdominance. Consequently, these data suggest that cases in which natural selection is presented with a phenotype that could not be predicted by one or both alleles at each locus, is a comparatively rare phenomenon.

References

Agulnik, S., Agulnik, A., & Ruvinsky, A. (1990). Meiotic drive in female mice heterozygous for the HSR inserts on chromosome 1. *Genetic Research, 55*, 97-100.

Batzer, M., & Deininger, P. (2002). Mammalian retroelements. *Genome Research, 10,* 1455-1465.

Chong, S., Youngson, N., & Whitelaw E. (2007). Heritable germline epimutation is not the same as transgenerational epigenetic inheritance. *Nature Genetics, 39*, 574-575.

Crick, F., & Orgel, L. (1980). Selfish DNA, the ultimate parasite. *Nature, 284,* 604-607.

Cui, X., Affourtit, J., Shockley, K., Woo, Y., & Churchill, G. (2006). Inheritance patterns of transcript levels in F1 hybrid mice. *Genetics, 174,* 627-637.

Dawkins, R. (1976). *The selfish gene*. New York: Galaxy Books.

de La Casa-Esperón, E., et al. (2002). X chromosome effect on maternal recombination and meiotic drive in the mouse. *Genetics, 161,* 1651-1659.

Dutrillaux, B. (1980). Chromosomal evolution of the great apes and man. *Journal of Reproduction and Fertility, 28,* 105-111.

Ensembl. (2007). Release 48. Available at: http://www.ensembl.org/Homo_ sapiens/index.html.

Gropp, A., & Winking, H. (1981). Robertsonian translocations: Cytology, meiosis, segregation patterns and biological consequences of heterozygosity. *Symposia of the Zoological Society of London, 47,* 141-181.

Hamerton, J., Canning, N., Ray, M., & Smith, S. (1975). A cytogenetic survey of 14,069 newborn infants. I. Incidence of chromosome abnormalities. *Clinical Genetics, 8,* 223-243.

Hassold, T., & Hunt, P. (2001). To err (meiotically) is human: The genesis of human aneuploidy. *Nature Reviews Genetics, 2,* 280-291.

Henikoff, S., & Malik, H. (2002). Centromeres: Selfish drivers. *Nature, 417,* 227.

Jorde, L., Carey, J., Bamshad, M., & White, R. (2000). Selfish genes, the phenotype paradigm and genome evolution. In *Medical genetics* (pp. 249, Table 12-3). New York: Mosby, Inc.

Kong A., et al. (2004). Recombination rate and reproductive success in humans. *Nature Genetics, 36,* 1203-1206.

Kong A., et al. (2008). Sequence variants in the RNF212 gene associate with genomewide recombination rate. *Science, 319,* 1398-1401.

Liu, Y., et al. (2006). Genetic linkage of human height is confirmed to 9q22 and Xq24. *Human Genetics, 119,* 295-304.

Pardo-Manuel de Villena, F., & Sapienza, C. (2001a). Nonrandom segregation during meiosis: The unfairness of females. *Mammalian Genome, 12,* 331-339.

Pardo-Manuel de Villena, F., & Sapienza C. (2001b). Female meiosis drives karyotypic evolution in mammals. *Genetics, 159,* 1179-1189.

Pardo-Manuel de Villena, F., de La Casa-Esperon, E., Briscoe, T., & Sapienza, C. (2000). A genetic test to determine the origin of maternal transmission distortion: meiotic drive at Om. *Genetics, 154,* 333–342.

Perola, M., et al. (2007). Combined genome scans for body stature in 6,602 European twins: evidence for common Caucasian loci. *PLoS Genetics, 3,* e97.

Rakyan, V., et al. (2003). Transgenerational inheritance of epigenetic states at the

murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 2538-2543.

Sapienza, C. & Doolittle, W. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature, 284*, 601-603.

Scherthan, H. (2007). Chromosome number in mammals. *Encyclopedia of Life Sciences*. Available at: http://mrw.interscience.wiley.com/emrw /9780470015902 /els/article/a0005799/current/pdf.

Suskind, P. (1986). *Perfume: The story of a murderer*. New York: Alfred A. Knopf.

Wu, G., et al. (2005). Maternal transmission ratio distortion at the mouse Om locus results from meiotic drive at the second meiotic division. *Genetics, 170*, 327-334.