Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems

Patrick C. Phillips
Center for Ecology and Evolution, University of Oregon, Eugene, OR 97403 USA and Instituto Gulbenkian de Ciência, P-2781-901 Oeiras Portugal. Email: pphil@uoregon.edu

Abstract

Epistasis, or interactions between genes, has long been recognized to be fundamentally important to understanding both the structure and function of genetic pathways and the evolutionary dynamics of complex genetic systems. With the advent of high throughput functional genomics and the emergence of systems approaches to biology, as well as a newfound ability to pursue the genetic basis of evolution down to specific molecular changes, there is a renewed appreciation both for the importance of studying gene interactions and for addressing these questions in a unified, quantitative manner.

To what extent can we understand the function and evolution of genetic systems by examining one gene at a time and to what extent do we need to worry about the potentially daunting number of possible interactions among the thousands or tens of thousands of genes operating within most organisms? Individual components or entire systems? Although this is an old topic

Web Summary

- Epistasis has been used to describe a number of phenomena, including the functional interaction between genes, the genetic outcome of mutations acting within the same genetic pathway, and the statistical deviation from additive gene action. Converging interests across genetics suggests that it is now a good time to develop a more unified view of epistasis and gene interactions.

- One of the traditional uses of epistasis analysis has been to order genes within developmental and metabolic pathways. These approaches have recently become much more systematic via the use of high throughput genetic screens, especially in yeast. These studies show that gene interactions are ubiquitous and can be used to help understand the structure of complex genetic networks.

- The major limitation of comprehensive analyses of gene interactions is the total number of interactions that must ultimately be tested, which grows at roughly the square of the number of genes (e.g., >18 million interactions in the yeast genome). Future work in this area will need to focus on particular subsets of this interaction space using information from other sources, such as functional genomics.

- Epistasis can be a major barrier to inferring the genetic basis of complex traits within natural populations. The effects of many quantitative trait loci (QTL) may be obscured by interactions with other loci, which can make mapping difficult.

- Human genetic disease is one area in which epistasis appears to be fairly common, although we have few examples in which the functional basis of a particular interaction has been demonstrated. Epistasis is one possible explanation for why human mapping results can be difficult to replicate.

- Epistasis arises as a natural byproduct of the evolutionary process, as all subsequent evolutionary change is built upon genetic changes that have occurred previously.

- There is clear evidence that epistasis helps to structure the possible pathways that evolution may follow, but even after nearly a century of debate on the topic, we still do not know if epistasis creates a major barrier to evolutionary change.

- An increased focus on the quantitative effects of gene interactions will provide the basis for a unified approach to studies of gene interaction, while providing a point of articulation between genetics, functional genomics, evolutionary genetics, and systems biology.

of debate within genetics, the recent rush towards combining comprehensive functional genomics and systems biology with high resolution genetic mapping is now providing the necessary empirical muscle to address these issues much more thoroughly than possible in the past. At the same time, a deeper understanding of the functional basis of gene interactions is generating an exciting intersection among a wide set of genetic disciplines, ranging from protein biochemistry to evolutionary genetics. We have never been in a better position to assess the role that gene interactions play within biological systems.

It has been roughly 100 years since William Bateson invented the term "epistasis" to describe the discrepancy between the prediction of segregation ratios based on the action of individual genes and the actual outcome of a dihybrid cross[1]. The usage of epistasis has since expanded to describe nearly any set of complex interactions among genetic loci (Box 1). Over the years geneticists have used epistasis to describe three distinct things: the functional relationship between genes, genetic ordering of regulatory pathways, and quantitative differences of allele-specific effects (Fig. 1). Using the same word to describe subtly different phenomenon has generated surprisingly little confusion in the literature—mostly because of a tendency for different areas of genetics to effectively ignore one another. This is no longer possible. Molecular geneticists are now studying how specific allelic effects traverse complex regulatory networks, while evolutionary geneticists are moving from statistical descriptions of genetic variation to identifying the specific nucleotide changes responsible for adaptive evolution. What has become clear in the century since the concept of epistasis was introduced, however, is that most of the systems that underlie cellular, developmental, and physiological function are composed of many elements that interact with one another in frequently complex ways. The challenges generated by the presence of epistasis provides a point for unification of traditionally disparate areas of research and shows that this fundamental genetic concept is more relevant now than ever.

Here, I first review various definitions of epistasis and show what they share in common and how they differ. I then turn to how the analysis of epistatic interactions between genes can be used to tease apart the global structure of these systems, and examine the impact that epistasis has on our ability to understand the genetic basis of natural variation, especially as it pertains to genetic variation associated with disease within human populations. I finally review various models of how evolution builds complex systems and explore how recent studies of molecular evolution can be used to determine the role that epistasis plays in directing the path of evolutionary change. A review as broad as this one naturally can not hope to cover all of the relevant literature. Luckily, there have been a number of recent, more specialized reviews that cover topics such as the evolutionary impacts of epistasis[2-5], the role of epistasis in complex traits[6-8], the impact of epistasis on human disease[9-11], statistical issues in detecting epistasis[12-17], and the use of synthetic interactions to define complex interaction networks[18,19].

Differing perspectives on gene interactions

Over the years the disparate needs of geneticists have lead to a plethora of different nuanced meanings for the term "epistasis", all of which involve gene interactions at various scales (Box 1). Although a few have suggested using the more generic term "gene interaction" to encompass the variety of phenomena labeled as epistasis so that epistasis can retain it original, more specialized meaning[1], given the history of use, this seems untenable. As will be evident below, traditional uses of epistasis to order genes within pathways have become increasingly quantitative, further obscuring boundaries of usage, and suggesting that expansion, rather than contraction of the usage of "epistasis" is the order of the day.

The uses of "epistasis" can be boiled down into three major categories, each of which have been called many things (so I will impose my own labels; Fig. 1, Box 1):

1. *Functional epistasis*. This term addresses the molecular interactions that proteins (and other genetic elements) have with one another, whether they operate within the same pathway or consist of proteins who directly complex with one another[18]. This is a strictly functional description, without a direct genetic link, although we would obviously predict a genetic consequence if the functional relationship were to be disrupted. Since it seems that the usage of epistasis is confusing enough already without it being adopted as a completely general description of complex systems, we would do well to avoid this usage and reserve epistasis solely to describe the consequences of allelic substitution, as described below. There are other descriptive terms, such as protein-protein interaction, that can be used in this context.

2. *Compositional epistasis*. This is a new term intended to describe the traditional usage of epistasis as the blocking of one allelic effect by an allele at another locus. The only way that this effect can be realized, however, is by combinatorially substituting one allele for another against a standard background. Thus, the genotypic composition of an individual is changed only at the loci of interest, with the rest of the background being invariant. Compositional epistasis therefore describes the way that a specific genotype is put together, or composed, and the influence that this specific genetic background has on the effects of a given set of alleles. If we expand this to include genetic interactions beyond those exposed in the double mutant homozygote, then we essentially have the definition of epistasis being used in modern systems biology. This does not have to be limited to qualitative categories of phenotypes, but includes quantitative measures of interaction against a fixed background as well (Box 2).

3. *Statistical epistasis*. This is the usage due to Fisher (Box 1), in which the average deviation of combinations of alleles at different loci is estimated over all other genotypes present within a population. Fisher's approach solves two problems. First, it yields the appropriate measures for describing evolutionary change[20]. Second, there are far more genetic combinations possible within a population (even in bacteria) than can ever be realized, so there is no such thing as a standard genetic background within a natural population; only the set of backgrounds most likely to be encountered within a given population sample. A compositional epistasis approach can not be formally applied to natural populations because it is impossible to enumerate all possible genetic interactions for any real population. Statistical epistasis can be thought of as the deviation induced by simultaneously substituting two alleles at different loci within a randomly selected individual within the population, after taking into account what we would expect the effect of substituting each allele separately would be. It is important to note that statistical epistasis does not simply mean that phenotypes are measured quantitatively, but instead that they are sampled from a population as opposed to being intentionally constructed.

The last two views of epistasis are actually complementary to one another. Compositional epistasis measures the effects of allele substitution against a particular fixed genetic background, while statistical epistasis measures the average effect of allele substitution against the population average genetic background. Measures of statistical epistasis are dependent on genotype frequencies, which would seem to make them a bit ephemeral. However, Fisher would undoubtedly point out that any measure of interactions is dependent on the specific genetic context in which it is measured, so simply choosing a fixed genetic background to test for allelic effects is equivalent to setting allele frequencies at other loci to 1.0. In this way, compositional epistasis is arbitrary in its own way. There is something of a tradeoff between precision and generality here. Statistical measures often average over variable epistatic effects at many different loci, which may tend to cancel out one another[2,21].

The hierarchical structure of the relationship between compositional and statistical epistasis is very similar, say, to the way that a Punnett square can be seen to be a special case of the Hardy-Weinberg condition under specific mating conditions. Some authors have referred to substitution against a fixed background as "physiological"[22] or "functional"[20] epistasis, but I do not favor these terms in this context because they are still essentially statistical measures and do not really tell us anything about gene function in the way that, say, a molecular biologist would use the term.

In the end, different uses of epistasis can in fact be unified under a single perspective, using the view that epistasis measures bi-allelic substitution under different genetic backgrounds: fixed or average. It is as simple—and as complex—as that.

Epistasis as a tool: Dissecting regulatory pathways

One of the first characters studied by Bateson and Punnett[23] that revealed a pattern of epistasis was flower color in sweet peas. As illustrated in nearly every genetics textbook, it is possible to cross two colorless (white) flowers and recover purple flowers in the offspring. The non-Mendelian segregation ratios of the F2's in this cross (9:7) suggest that two complementary genes are interacting with one another (Fig. 2). Our modern interpretation is that these genes produce enzymes that operate within the anthocyanin pathway such that a mutation in either gene can disrupt flower color. In general, the fact that the phenotype of an individual depends strongly on the specific combination of alleles at two or more loci suggests immediately that this dependency must be saying something about the nature of the functional interaction between these loci (Fig. 1). Although many anecdotal cases were collected towards the beginning of the twentieth century, this idea was exploited most fully by Beadle and Tatum and their students during the advent of biochemical genetics[24]. Separate knockouts that each disrupt a particular function can be crossed together in order to observe the nature of the interaction and thereby order the genetic pathway[25]. This can be used most effectively in organisms in which large numbers of mutants can be generated, crossed and phenotyped[26]. For instance, in *C. elegans* epistasis analysis has been used extensively to order dozens of genes into pathways affecting diverse traits such as sex determination[27], the development of the vulva[28], and entry into the dauer resting stage[29]. In each case, ordering of the regulatory pathways genetically preceded molecular characterization and provided a strong set of functional models or hypotheses that could then (1) be used to make predictions about the probable functions of the identified genes and (2) be tested by the molecular characterization of the gene products.

High throughput approaches

The vast majority of studies that have used epistasis to analyze the structure of genetic pathways have used a small set of genes that had previously been identified to influence the trait of interest using single mutant analysis. However, the entire premise of epistasis is that genetic interactions can generate novel phenotypes when found in combination with one another. How better to discover such interactions than by looking for interactions between randomly selected genes? Even better, why not conduct a systematic study of the possible pair-wise interactions between all genes? The problem here, of course, is one of scale. The number of pair-wise interactions between genes grows at roughly the square of the number of genes [$n(n-1)$ to be exact, where n is the number of loci; $n(n+1)$ if the parental strains are included; half these numbers if reciprocal interactions do not need to be tested]. A daunting, but doable task for the 190 interactions resulting from twenty genes, but something nearing impossible for the over 18 million possible interactions for every gene in the yeast genome. Despite this challenge, such comprehensive approaches are beginning to be executed, spurred on by the availability of comprehensive deletion and/or knockdown libraries and high throughput maintenance and screening techniques.

Among the first studies of this type was that of Tong et al.[30], who used a high throughput analysis approach (synthetic genetic array analysis) to examine the interactions of eight different deletion mutants against an array of ~4,700 other deletion backgrounds. They later expanded this query set to 132 different genes[31]. The interactions revealed here define a network of ~1000 genes involving ~4000 interactions, with most interactions tightly clumped within self-similar functional clusters. These results can go a long way towards helping us understand how complex gene networks can build robustness into cellular systems[32], but because they are based on a growth/no growth criterion, they are largely qualitative[33]. A nice example of how classical compositional epistasis can be combined with a more quantitative approach is St. Onge et al.'s[34] examination of the genetic interaction system influencing the resistance of yeast to the mutagen methyl methanesulfonate (MMS). Using 26 mutants known to influence MMS resistance, St. Onge et al. constructed all 650 possible double-deletion strains. Of these 10 interactions generated synthetic lethality, while 67 were classified as "aggrevating interactions" (negative epistasis; Box 2) and 45 were classified as "alleviating interactions" (positive epistasis) based on a multiplicative model. The interactions within this latter class, which provide greater than expected resistance to MMS, were used to generate a functional interaction map among the loci (Fig. 2). Most known genetic pathways were recovered, in addition to a few novel connections. Interestingly, nine of the ten deletion combinations that yielded essentially the same growth characteristics both in the single and double mutants (a "coequal" relationship) involved direct interactions between protein subunits, suggesting that it might be possible to connect particular epistasic outcomes with specific forms of functional interaction.

More complex synthetic interactions

There is no reason to expect all forms of epistasis to be revealed simply by the absence of a gene (certainly a sledgehammer approach for perturbing complex systems). For example, Kroll et al.[35] devised a method for looking for interactions induced after systematically overexpressing genes. Using this approach, Sopko et al.[36] found that about 15% of a set of 5280 yeast genes show a growth defect when overexpressed, with most of the overexpression effects not matching the phenotypes of their corresponding deletions. Testing a deletion of the cyclin-dependent kinase *pho85* over the overexpression library revealed 65 synthetic interactions, most of which were previously unknown. There are of course an endless combination of knockouts, overexpression, natural, and induced alleles that can be used in combination to probe a given library or array. The fact that the specific interaction results obtained depend on the nature of the "probe" (deletion, overexpression, etc) in the few studies that have been conducted thus far is perhaps not surprising, but does indicate that the overall structure of the interaction network is likely to be complex and allele-specific. There is also the possibility of something like a "network uncertainty principle": perturbing one part of the network is likely to cause changes in the nature of the interactions between other elements of the network[37].

Synthetic mapping via RNAi knockdowns

The existence of comprehensive deletion libraries and high throughput screening methods have made yeast a particularly powerful system for systematically dissecting epistatic interaction networks (reviews in Refs 18,[19]). Such approaches are unfortunately unlikely to be widely applicable, so what about other organisms? Again, the problem here is one of scale. How do we generate and score so many possible combinations? Perhaps the cleverest approach for getting around this problem in multicellular organisms has been the use of RNAi libraries to knockdown (rather than knockout) genes in a systematic fashion. Lehner et al.[38] used 37 strains of *C. elegans* with mutations in cell signaling components to "query" a RNAi library of ~1750 genes involved in signal transduction, transcriptional regulation and chromatin remodeling. The query was executed by raising each strain on bacteria expressing double stranded RNA of

the target gene of interest and observing how many of these ~65,000 combinations resulted in disrupted growth and/or reproduction. This yielded a genetic network of 349 interactions involving 162 genes. While the number of interactions may seem low versus the total number possible, it is on the same order as that observed in yeast, which displays around 0.6% of interactions for non-essential genes[31], with the frequency of interactions for essential genes being much higher at 3.3%[39]. Despite these gross similarities, it appears that the structure of the interaction networks between yeast and worms is quite divergent, with perhaps fewer than 5% of the interactions shared in common[40].

Integrating epistasis data with data from other sources

The obvious next step for these analyses is to link the network structure revealed by epistasis analysis to information obtained from other methods, such as yeast two-hybrid, chromatin immunoprecipitation, and gene expression assays to build a comprehensive map of the full interactome[41-44]. Because the scale of experiments required, a fruitful approach for now appears to be to concentrate on a large, but finite set of genes known to be involved in a well defined biological process. For example, Collins et al.[45] used all pair-wise interactions of 743 genes known to influence chromosomal processes such as DNA repair, transcriptional regulation, and chromatid segregation in yeast. This allowed them to place their interaction results within the context of already well-explored systems, such as the biochemistry of the multiprotein Mediator transcriptional coactivation complex. In this case, epistatic interactions corresponded well to known physical interactions among proteins, yet allowed novel interactions to be detected above what would otherwise be a chaotic set of more than half a million potential interactions.

A continuing challenge for the future is figuring out ways of overcoming the inherent scaling problem of the exponential growth in the number of possible genetic interactions. Further, now that we are beginning to look at pair-wise interactions, how do we know that third or higher order interactions will not also be relevant[20]? Returning to the theme of this section, however, high throughput screens by themselves are not going to explain the basis for why the interaction exists in the first place. Interaction networks are simply hypotheses that need to be rigorously tested using other functional approaches. In this, epistasis analysis has proven to be a valuable tool whose use is sure to continue to grow.

Epistasis as an obstacle: Genetic mapping of complex traits and diseases

The presence of epistasis can greatly obscure the mapping between genotype and phenotype. In contrast to mutation based studies, which start with a known genetic lesion and then ask how a specific locus interacts with other loci, the goal of complex trait analysis, or quantitative trait locus (QTL) mapping, is to start with a given set of phenotypes and to then identify the genes responsible for generating differences among individuals within a population. Some of the issues that arise in QTL mapping in the presence of gene interactions are nicely illustrated by a recent study of the genetic basis of the response to selection on body size in chickens by Carlborg et al.[46] (Fig. 3). After roughly 40 generations of selection, males in the low weight line weighed six times less than males in the high weight line. There must surely be a strong genetic signal in such a difference. Yet, after examining many marker loci for their individual effects, only one QTL (prosaically named Growth9) appeared to have an effect, and the signal for that was itself weak. However, by looking for epistatic relationships among the markers, the authors were able to identify five additional genomic regions with significant effects on growth, each of which only showed their effects when on the high-growth Growth9 background (Fig. 3). Together, this loose network of epistatic genes accounted for 45% of the difference among the selected lines, an overall effect of 3.3 phenotypic standard deviations. The "individual" effect of Growth9 was in fact completely accounted for via its epistatic interactions with the other QTL. A very similar pattern of modules of interacting QTL have

been identified as influencing obesity in mice[47]. Although the genes underlying these QTL still need to be identified, it is clear that the vast majority of the genetic information in this system would have been missed without also looking for possible interactions.

Similar hidden effects are undoubtedly lurking within natural populations as well. For example, Ehrenreich et al.[48] used association mapping at 36 candidate loci to investigate the genetic basis of natural variation in shoot branch architecture within populations of *Arabidopsis thaliana*. They were able to identify three loci with significant associations with morphology in the wild, but none of these loci were implicated in a standard QTL mapping experiment involving recombinant inbred lines. Interestingly, however, these loci did exhibit significant epistatic relationships among one another within these lines. The authors conclude that epistasis may be prevalent within these populations. If nothing else, this study illustrates that moving between association and QTL mapping studies can be complicated by genotype-specific patterns of epistasis. Association mapping in natural populations will be based on statistical epistasis whereas QTL mapping in between two inbred lines draws closer to compositional epistasis[49,50] (although the total number of segregating backgrounds is still huge). Similar interactions underlying complex traits have been found in odor sensing behavior in Drosophila[51], growth and yield in tomatoes[52], the *Arabidopsis* metabolome[53], the skeletal architecture of mice[54], and in large scale studies of yeast growth[55], morphology[56], and gene expression[57,58].

Variation within natural populations

From mutational studies, we know that epistasis in the classical sense is ubiquitous because genes interact in hierarchical systems to generate biological function. For quantitative genetics and the genetics of complex traits however, it is not the total scaffold of biological function that matters, but the residual variation segregating within natural populations that determines differences among individuals. Traditionally, quantitative genetics has focused on aggregate measures, such as genetic variance and heritability, to estimate genetic effects. There seems to be little evidence that epistatic variance plays an important role in most populations[59], although epistasis at individual loci can make significant contributions to additive variance[60]. Now that we are beginning to be able to dissect the specific genetic basis of complex traits, will epistasis have a larger role to play? The answer is that we still do not really know. This is partially because, even after several decades of work in this area, identifying the causal basis of individual variation in complex traits has remained fairly elusive and partially because the statistical issues involved in estimating large numbers of potential interaction effects has limited the power of most existing studies. Both of these barriers are beginning to fall.

Epistasis in human health and disease

By its very nature, epistasis is a property of whole genotypes. Epistatic effects are therefore most clearly revealed in the eccentricities of particular individuals. It is perhaps therefore not surprising that some of the best examples of epistasis are emerging from an area in which the focus on the individual reigns above all else: human health. Here we have the most complex of complex traits. Part of the motivation for the recent enthusiasm for looking for genetic interactions underlying human disease is the sense that previous failures to identify, and especially to replicate, significant individual genetic effects might be driven by underlying complexity generated by epistasis[6,11]. Indeed, epistatic shielding of disease alleles is one possible explanation for their persistence within populations[61]. Given the rapid increase in the size and precision of human association studies, we are now entering an era in which we can rigorously address the hypothesis of whether previous problems are a function of limitations in the data or truly the result of genetic complexity[62].

There are numerous cases of epistasis appearing as a statistical feature of association studies of human disease. A few recent examples include coronary artery disease[63], diabetes[64], bipolar effective disorder[65], and autism[66]. Unfortunately, in only a few cases has the functional basis of these potential interactions been revealed. One interesting exception involves genetic interactions underlying the autoimmune disease multiple sclerosis. Here, Gregersen et al.[67] found evidence that natural selection might be maintaining linkage disequilibrium between several different histocompatibility loci (*DR2a* and *DR2b*; Fig. 3) known to be associated with multiple sclerosis. Linkage disequilibrium can be generated by strong epistasis among adjacent loci (Box 2). To test this idea, Gregersen et al. genetically engineered mice to produce the appropriate human immune proteins and found that mice producing the *DR2b* protein were highly susceptible to disease, whereas those producing *DR2a* did not progress towards disease at all. Then in the critical test, mice expressing both genes had an overall reduced susceptibility to disease, suggesting that *DR2a* modulates the impact of *DR2b*. One possible model for this interaction is that *DR2b* stimulates the production of T-cells sensitive to the antigen that ends up inducing MS, whereas *DR2a* suppresses or even leads to the death of these cells[68] (Fig. 3). Such an interaction could help to explain why these negative effects could be segregating within human populations in the first place: under most conditions the influence of these two factors balance against each other, presumably to generate a heightened response to real pathogens. Multiple sclerosis is a complex disease with a fairly weak genetic signal[69], and the epistatic effect of these two immune genes has yet to be tested in humans because natural recombinants between *DR2a* and *DR2b* have yet to be observed[68]. These issues only serve to highlight how difficult it can be to identify underlying complex diseases, let alone the extra complications that can arise when epistasis between two loci dramatically affects disease penetrance.

Limitations to inference

As the field moves toward whole genome association mapping, the problem of scale that pervades all interaction tests becomes intense. The total number of tests required suggests that very stringent significance thresholds will be needed to control against false positives, but this in turn means that the only epistatic effects that will be detected will have to be really gigantic and/or sample sizes will need to be very large[70]. Indeed larger studies are more prone to detect epistasis than smaller ones[6]. Suggestions to limit testing to only those QTL with significant main effects[71] are probably ill-advised because (1) epistatic interactions with the largest relative effect sizes will be those with small main effects and (2) we already know that epistasis is frequently detected in the absence of main effects[50,72]. Carlborg and Haley[6] advocate a sequential approach where potentially interesting QTL are first identified using a high false discovery rate, and then used for subsequent tests for genetic interactions. Specialized breeding designs can be used to increase the resolution and ability to detect complex interactions[13, 73]. As above, the most productive approaches are probably going to involve coupling mapping strategies with other functional assays in an effort to focus on the interactions that are most likely to matter in practice[42,43]. While there is certainly strong evidence that epistasis can be important in determining variation in natural and human populations, only further detailed studies will tell us whether this is a widespread or limited phenomenon.

The structure and evolution of complex systems

Where does all of this epistasis come from in the first place? Is there something about the evolution of genetic systems that yields epistasis as a by-product? Because evolutionary change is predicated on the current state of a genetic system, functional epistasis is in fact an extremely likely outcome of the evolutionary process. Since future changes are built upon past changes, the "tinkering" nature of evolution[74] has the potential to build somewhat baroque systems. As solutions to one functional problem become fixed within an evolutionary lineage, future

functional changes will frequently be built by adding additional elements to these existing systems, as for example when new effector molecules attach themselves to the backbone of an existing signal transduction pathway. This will be true whether or not epistatic variation is present or important while selection is operating (Fig. 4).

Under this view, evolving genetic systems are something of a house of cards. Removing one central component can bring the whole house down (i.e., be epistatic to many other genes) more because of the overall structural dependence induced by historical contingency than because it is a result of some intricately pieced together machine[75]. Indeed, Crow[76] has conjectured that alleles with more severe effects, such as knockouts, will be more likely to display epistasis than alleles with more subtle genetic effects because larger perturbations are more likely to disrupt the overall structure of the genetic system. Thus, the fact that perturbation approaches, as outlined above, commonly reveal epistasis does not necessarily mean that the alleles responsible for evolutionary change also tend to be epistatic. Each allelic difference, including those generated via induced mutations, needs to be evaluated in its own light. Although epistasis is usually portrayed as a property of a given locus, as we have seen, it is really a property of individual alleles at multiple loci. Unfortunately, allelic variation for epistatic effects has yet to be studied in a systematic fashion[2].

It is important to remember that most models of the evolution of genetic systems (such as those depicted in Figure 4) represent very simple metaphors of complex genetic phenomena. One of the problems in this approach is representing complex, multidimensional processes as three-dimensional cartoons. It is clear that taking these kinds of cartoons too literally can lead to a limited view of possible evolutionary dynamics, such as neglecting the possibility of complex ridges connecting regions of high fitness. Fisher's[77] view was the evolutionary process was so multidimensional that there will always be some axis along which selection can move a population, such that adaptive valleys, even if they exist, will be very localized in their impacts. Kauffman[78] has emphasized the opposite, showing that the number of valleys can rapidly increase with increasing dimensions. To a large extent, this is an empirical question—albeit one that is extremely difficult to address adequately.

The bottom line here is that epistasis and genetic interactions are an inevitable consequence of the evolutionary process, no matter how it is conceived. This means that functional biologists have to confront the reality of complex genetic systems no matter what their ultimate cause. This is the exquisite—and sometimes frustrating—result of 3.5 billion years of descent with modification.

Epistasis and the path of evolutionary change

Epistasis can have an important influence on a number of evolutionary phenomena, including the genetic divergence between species[79], the evolution of sexual reproduction[4], and the evolution of the structure of genetic systems[80]. One of the more interesting long term questions in evolutionary biology is whether or not epistasis determines the path of evolutionary change. Although the focus here has traditionally been on interactions between disparate loci, currently the best systems for investigating this question are derived from functional studies of interactions operating within individual proteins (Box 3). Thus far, these studies[81-85] have shown that epistasis can play a strong role in limiting the possible paths that evolution can take, but not in limiting its eventual outcome. Of course this might be due in part to the fact that inaccessible evolutionary outcomes may never be observed, but this in itself is an important result. These studies have been especially valuable in helping to build a bridge between the functional analysis of epistasis that has characterized molecular genetics and the long term impact of epistasis on genetic change that has characterized much of the debate in evolutionary biology.

The evolution of regulatory complexity

One consequence of a systematic search for gene interactions is that the consequences of linkage may tend to be overlooked. As seen in the case of the histocompatibility loci in multiple sclerosis, linkage can facilitate the maintenance of epistatic interactions (and visa versa)[86] and could help explain how molecular complexity evolves. Such linkage is self evident when looking at evolution of protein function, but recent analysis of patterns of gene regulation suggest that there can be very complex patterns of gene regulation within localized genomic regions[87] that may be the result of similar types of evolutionary constraints. We need to look at interactions between promoters and coding genes, micro RNAs, chromatin remodeling, and other factors that Bateson would never have dreamed of, as being parts of epistatic networks whose evolutionary dynamics may be guided by complex sets of genetic interactions and their genomic relationship with one another.

Conclusion: Building toward the future

It should be apparent that the global analysis of gene interaction patterns bears a striking resemblance to what is now called "systems biology"[88]. One of the central questions in this field is whether there are emergent properties of complex systems that are not predicted from looking at individual system components, yet are essential for understanding the function of the system as a whole. From an evolutionary standpoint, we might also add questions such as whether the structure of the system has evolved to facilitate these properties (e.g., robustness, modularity, evolvability[33,80,89]). The answers to these questions will rely on our ability to expand the use of epistasis in two directions. First, as has already been occurring in a few model systems, we need to explore more of the potential interaction space through high throughput screens of genetic interactions, transcriptional regulation, protein modification and interaction, and phenotypes. Second, we need to complete the unification of classical and statistical views of gene interaction by encouraging molecular biologists to continue to become more quantitative in their measures of genetic outcomes and evolutionary geneticists to become more mechanistic in their interpretations of evolutionary change. As this occurs, all sides of epistasis (Figure 1) should become unified through the metaphor of quantitative flow across a genetic network. This approach can be used to predict the emergence of epistasis in the traditional sense[90], can facilitate the use of knockout and gain of function studies to test system-level predictions[91-93], and can help direct tests that should lead to the functional nature of the interactions themselves. This quantitative detail can then be used to understand the implications of these interactions from systems and evolutionary viewpoints in order to understand the broader population level consequences of epistasis for generating differences among individuals. This is neither reductionistic nor holistic, but a powerful combination of the two. The overwhelming combinatorics of the problem is a major issue (and at some point insurmountable), so progress will ultimately need to be based on strong hypotheses generated from functional information. Given recent work in this area, it is likely that for the next 100 years the concept of epistasis will be even more central to biology than it has over the last 100.

Box 1

Epistasis: what's in a name?

There have been many different uses of the term "epistasis" over the last 100 years, which leads to the potential for some confusion now that more biologists from different areas of genetics are increasingly looking at gene interactions. The original definition comes from William Bateson[94], who was specifically concerned with the observation that in some dihybrid crosses, not all possible phenotypic classes were be observed and/or that some gene combinations resulted in novel phenotypes. Some of the mutations seemed to be "stopping" or "standing above" the effects of other mutations. Such mutations were said to be *epistatic* (the ones being blocked, *hypostatic*). It was clear from these circumstances that

the mutations must be interacting with one another, at least in the loose sense that they exist within pathways that both influence the same phenotype. It was therefore perhaps natural that R.A. Fisher[95] used a derivative of this term, "epistacy" to mean any statistical deviation from the additive combination of two loci in their effects on a phenotype (Box 2). Unfortunately, population geneticists rapidly adopted the term "epistasis" to apply to this second, much more general class of phenomena[1], and so we are left with a situation in which geneticists studying genetic segregation of (usually) discrete phenotypes mean one thing by epistasis, whereas population and quantitative geneticists mean something slightly different. It is especially troubling that finding epistasis in one context (say during segregation in a specific cross) does not necessarily mean that there will be epistasis in the other context (say in the statistical sense). Worst of all, the opposite will frequently be true: an absence of the detection of epistasis in the statistical sense does not mean that there are no interesting interactions between loci in the stricter genetic sense[12,88]. There is perhaps a bit of irony in the fact that most scientists who work with epistasis rely on context to define the type of gene interaction that they are referring to.

There are a few other conceptual barriers to generating a more unified approach to gene interactions and epistasis. One issue arises from the way that one views how organisms are assembled. Are organisms constructed, with genes and their individual effects serving as the building blocks, or do organisms come to us as wholes, with each component only being understandable within the context of the system as a whole[96]? The former characterizes the approach to epistasis followed by most population geneticists, who tend to build up genotypes as though each allele has a specific predetermined effect that can be perturbed in certain circumstances by an interaction term describing epistasis. This viewpoint also fits in well with a traditional mutational approach to examine the function of a gene, since the mutant and wildtype functions of genes can be examined, manipulated and combined. The second approach has been used by quantitative geneticists and others studying natural variation or complex allelic series because in this context it become unclear what is part and what is whole—what is the reference standard against which each allele can be tested? Since there is no such thing as a naked gene without some broader genomic context, the building block model must ultimately be left to the theoreticians, but any real data will have to be examined using an effects model (Box 2).

Box 2

Measures of epistasis

In principle, detecting epistasis using the Bateson's definition is relatively straightforward because the phenotypes are qualitative and few in number. Once epistasis is made more quantitative and is expanded to include nearly any kind of genetic interaction, then things get a bit more complex. First of all, epistasis means that something different happens when a particular set of alleles from different loci are in combination than when apart. But different from what? It must be different from what we would expect if the effects of the two loci were combined independently. Here, however, the scale of measurement becomes important. Fisher defined epistasis as a deviation from the additive expectation of allelic effects[95]. For a haploid model, this might look something like $W_{xy} = \alpha_x + \alpha_y + \varepsilon$, where α is the individual effects of each allele at loci x and y and ε is the deviation due to epistasis. Relationships for diploids are more complex because of the possibility of interlocus interactions with the dominance state of the other locus[14]. Fisher presumably chose this definition because additive linear models are very tractable from a statistical point of view. Part of the reason that Fisher did not think that epistasis was that important is because he felt that there would usually be some scale to which the phenotypic values could be transformed such that the effects would be additive.

In the late 1960's, population geneticists started using deviation from a multiplicative model of gene action as the definition of epistasis instead of the additive model. This is because the evolutionary trajectories of loci with multiplicative fitness are independent of one another. In particular, if no linkage disequilibrium is present in the ancestral population, then none will develop if fitness effects are multiplicative[86]. A multiplicative haploid model would look something like $W_{xy} = \alpha_x \alpha_y + \varepsilon$. It is a little appreciated fact, however, that if linkage disequilibrium is already present in the base population, then it can still be maintained under a multiplicative model[97], so there is in fact no perfect scale with which to measure epistasis. For some traits, such as fertility, an additive scale might be natural, whereas for other traits, such as mortality, the multiplicative approach is probably more appropriate. Perhaps not surprisingly, different measures can lead to different interpretations of epistasis[98]. Aylor and Zeng[99] discuss possible extensions to common models of epistasis that attempt to span classical and statistical frameworks.

The fact that linkage disequilibrium can be generated by epistasis is sometimes proposed as an indicator of gene interaction, although this would frequently be expected to be a weak effect relative to other factors such as admixture, and, as we see, this expectation depends on how one measures epistasis. The stability of linkage disequilbrium depends strongly on the recombination rate, especially when linkage is tight[97]. Now that we have a firm idea of genomic structure of many organisms, we now know that there can often be very many genes with recombination map distances less than 0.01 or 0.05, so this might not be a trivial effect for many genes.

Much like the definitions of epistasis itself, there is a virtual menagerie of terms associated with particular forms of epistatic effects. Examples include: synergistic, diminishing, antagonistic, aggravating, ameliorating, buffering, compensatory, and reinforcing. Most of these refer to similar phenomenon, which makes it difficult to keep things straight. Take for example synergistic epistasis. This occurs when an individual with a particular two-locus combination of alleles displays a phenotype beyond that expected from the individual effects of the alleles. If these are, say, deleterious mutations, then that means that the phenotype is less than expected, but for positive mutations this means that the phenotype is greater than expected. So sometimes synergistic means "extra good" and sometimes it means "extra bad". The field would be paid a great service if all of these redundant names, whose meaning depends strongly on context, were done away with in favor of two simple terms: *positive* and *negative epistasis*[100]. Positive epistasis means that the phenotype is higher than expected and negative epistasis means that the phenotype is lower than expected. These terms are preferable because (1) their meaning is immediately apparent with minimal thought and (2) it turns out that it is the sign of the epistasis that matters in most evolutionary processes (such as the generation of linkage disequilbrium), not the change in relative direction of the effects of the individual loci[2]. The latter relationship can be addressed using another simple term, *sign epistasis*[5]. This indicates that the direction of the epistatic and individual effects differ from one another that the direction of selection on the individual alleles actually changes depending on the genetic context. For example, if a pair of mutations lower fitness when found individually, but increase fitness when found together, then this results in an adaptive valley, which has very different functional and evolutionary implications than if the sign of the individual effects and epistasis are in the same direction (Fig. 4).

Box 3

Epistasis within a locus

One of the best systems for rigorously testing the functional and evolutionary consequences of epistasis is in the within-locus interactions that characterize protein folding and activity.

The nicest example of this thus far comes from Ortlund et al.'s[82] investigation of the evolution of novel function in vertebrate steroid receptors. The first step here was to use phylogenetic methods to reconstruct the inferred ancestral protein sequence that predates the separate evolution of the mineralcorticoid and glucocorticoid steroid receptors and to test its function[81]. It turns out that the ancestral protein is fairly promiscuous and interacts with a variety of steroid ligands, even with ligands not present within the ancestral organism at that time. Specialization therefore occurred via the evolution of a glucocorticoid-specific receptor from a more general mineralcorticoid ancestor, which was achieved via changes at two interacting sites, S106P and L111Q. By itself, S106P essentially destroys receptor function, while L111Q by itself has little functional effect. Together, however, S106P changes the architecture of the protein in such a way that allows L111Q move down to form a novel hydrogen bond with cortisol, which is a clear case of functional epistasis (**a**). Ortlund et al. call these two changes "group X". Three more amino acid changes (group Y) are needed to yield the final specificity to cortisol, but these substitutions destabilize the protein. They must therefore be preceded by two other amino acid changes (group Z) that stable the purturbation in protein structure induced by changes in the X and Y groups. Ortlund et al. call the Z group substitutions "permissive" mutations, because they appear to have little effect on receptor function, but are a critical first step for allowing the other functional changes to occur. There is another permissive mutation, Y27R (**b**), which precedes all of these other changes and which generates a novel cation-π interaction (replacing a weaker hydrogen bond) that stabilizes portions of the structure that would have otherwise been destabilized by the subsequent changes.

Together, these structural interactions create a specific order in which the evolutionary substitutions must occur. There are a number of possible pathways for these changes (**c**), but only a few are functionally viable because the so-called "conformational epistasis" generated by structural failure of the protein limits the evolutionary options. Here the evolution is from a generalized response in the ancestor (AncGR1) to the hormones aldosterone (green), cortisone (purple) and DOC (orange) to specificity to cortisone alone (+XYZ). In this example we have a direct tie between specific amino acid changes, epistatic interactions generated by their influence on protein structure, and the impact that these interactions have on subsequent evolutionary change. Figure reprinted with permission from Ref 82.
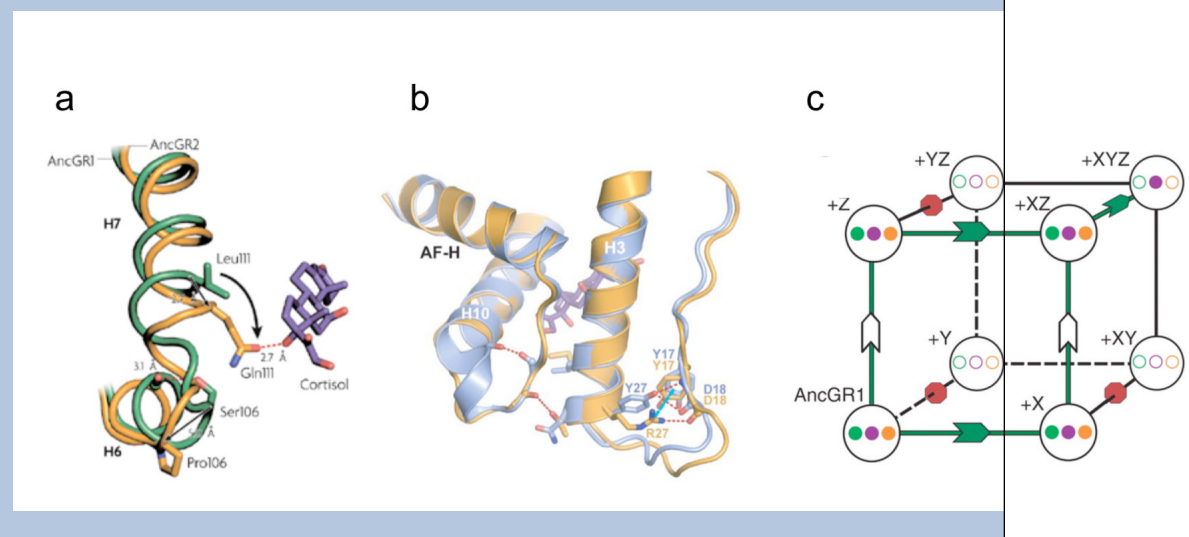
Acknowledgements

About the author

Patrick Phillips received his Ph.D. in evolutionary biology from the University of Chicago and did his postdoctoral work with James Crow in the Laboratory of Genetics at the University of Wisconsin, Madision. He is currently a professor of biology in the Center for Ecology and Evolutionary Biology and the Department of Biology at the University of Oregon, where his lab focuses on theoretical and empirical studies of complex traits, using the nematode *C. elegans* and it relatives to pursue the molecular genetics of the genotype-phenotype map for traits such as reproductive success, sexual interactions, longevity, and the behavioral response to temperature and chemicals.

Glossary terms

PUNNET SQUARE, A method of calculating the outcomes of a genetic cross by multiplying the expected frequency of gametes from a mother (rows) times the expected frequency of gametes from the father (columns); HARDY—WEINBERG EQUILIBRIUM, (HWE) A theoretical description of the relationship between genotype and allele frequencies that is based on expectation in a stable population undergoing random mating in the absence of selection, new mutations and gene flow; under these conditions (and in the absence of linkage disequilibria) the genotype frequences are are equal to the product of the allele frequencies; DAUER LARVA, A developmentally arrested, immature, long-lived and non-feeding form of *Caenorhabditis elegans* that forms under conditions of food scarcity and high population density, and that resumes development if food levels increase; SYNTHETIC LETHAL MUTATIONS, Two mutations are considered to be synthetically lethal if in combination they result in death, whereas an individual with either mutation alone is viable; CHROMATIN IMMUNOPRECIPITATION, (ChIP) A technique used to identify potential regulatory sequences by isolating soluble DNA chromatin extracts (complexes of DNA and protein) using antibodies that recognize specific DNA binding proteins; LINKAGE DISEQUILIBRIUM, (LD). A measure of whether alleles at two loci coexist in a population in a nonrandom fashion. Alleles that are in LD are found together on the same haplotype more often than would be expected under a random combination of alleles; ADMIXTURE, The pattern of genetic variation that results when a population is derived from founders that originated from more than one ancestral population.

References

1. Phillips PC. The language of gene interaction. Genetics 1998;149:1167–1171. [PubMed: 9649511]

2. Phillips, PC.; Otto, SP.; Whitlock, MC. Epistasis and the Evolutionary Process. Wolf, JD.; Brodie, ED., III; Wade, MJ., editors. Oxford University Press; Oxford, England: 2000.

3. Malmberg RL, Mauricio R. QTL-based evidence for the role of epistasis in evolution. Genet Res 2005;86:89–95. [PubMed: 16356282]

4. Otto SP, Gerstein AC. Why have sex? The population genetics of sex and recombination. Biochem Soc Trans 2006;34:519–522. [PubMed: 16856849]

5. Weinreich DM, Watson RA, Chao L. Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. Evolution 2005;59:1165–1174. [PubMed: 16050094]

6. Carlborg O, Haley CS. Epistasis: too often neglected in complex trait studies? Nat Rev Genet 2004;5:618–625. [PubMed: 15266344]

7. Holland JB. Genetic architecture of complex traits in plants. Curr Opin Plant Biol 2007;10:156–161. [PubMed: 17291822]

8. Wade MJ. Epistasis, complex traits, and mapping genes. Genetica 2001;112-113:59–69. [PubMed: 11838787]

9. Azevedo L, Suriano G, van Asch B, Harding RM, Amorim A. Epistatic interactions: how strong in disease and evolution? Trends Genet 2006;22:581–585. [PubMed: 16911842]

10. Nadeau JH. Modifier genes in mice and humans. Nat Rev Genet 2001;2:165–174. [PubMed: 11256068]

11. Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 2003;56:73–82. [PubMed: 14614241]

12. Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet 2002;11:2463–2468.2468 [PubMed: 12351582] A clear review of the limitations in moving from statistical estimates of epistatic effects to understanding genetic causation.

13. Demuth JP, Wade MJ. Experimental methods for measuring gene interactions. Annual Review of Ecology, Evolution and Systematics 2006;37:289–316.

14. Musani SK, et al. Detection of gene × gene interactions in genome-wide association studies of human population data. Hum Hered 2007;63:67–84. [PubMed: 17283436]

15. McKinney BA, Reif DM, Ritchie MD, Moore JH. Machine learning for detecting gene-gene interactions: a review. Appl Bioinformatics 2006;5:77–88. [PubMed: 16722772]

16. Marchini J, Donnelly P, Cardon LR. Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat Genet 2005;37:413–417. [PubMed: 15793588]

17. Alvarez-Castro JM, Le Rouzic A, Carlborg O. How to perform meaningful estimates of genetic effects. PLoS Genetics 2008;4:e1000062. [PubMed: 18451979]

18. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. Nat Rev Genet 2007;8:437–449.449 [PubMed: 17510664] A very comprehensive review of existing work on using high throughput approaches in yeast to dissect complex gene interaction networks. Includes a good discussion of the overall conceptual framework.

19. Costanzo M, Giaever G, Nislow C, Andrews B. Experimental approaches to identify genetic networks. Curr Opin Biotechnol 2006;17:472–480. [PubMed: 16962766]

20. Hansen TF, Wagner GP. Modeling genetic architecture: a multilinear theory of gene interaction. Theoretical population biology 2001;59:61–86. [PubMed: 11243929]

21. Elena SF, Lenski RE. Test of synergistic interactions among deleterious mutations in bacteria. Nature 1997;390:395–398.398 [PubMed: 9389477] Uses randomly generated mutations in *E. coli* to demonstrate that epistatic effects between loci can be highly variable and frequently cancel out one another.

22. Routman EJ, Cheverud JM. Gene effects on a quantitative trait: Two-locus epistatic effects measured at microsatellite markers and at estimated QTL. Evolution 1995;51:1654–1662.

23. Bateson, W.; Saunders, ER.; Punnett, RC.; Hurst, CC. Reports to the Evolution Committee of the Royal Society, Report II. Harrison and Sons; London: 1905.

24. Beadle GW. Genetics and metabolism in Neurospora. Physiological Reviews 1945;25:643–663.

25. Avery L, Wasserman S. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. Trends Genet 1992;8:312–316. [PubMed: 1365397]

26. Huang LS, Sternberg PW. Genetic dissection of developmental pathways. Worm Book. 2005 doi/ 10.1895/wormbook.1.88.2. http://www.wormbook.org. A comprehensive treatment of how to use classical epistasis analysis to reconstruct genetic pathways.

27. Goodwin EB, Ellis RE. Turning clustering loops: sex determination in *Caenorhabditis elegans*. Curr. Biol 2002;12:R111–R120. [PubMed: 11839294]

28. Sternberg PW, Horvitz HR. The combined action of two intercellular signaling pathways specifies three cell fates during vulval induction in *C. elegans*. Cell 1989;58:679–693. [PubMed: 2548732]

29. Thomas JH, Birnby DA, Vowels JJ. Evidence for parallel processing of sensory information controlling dauer formation in *Caenorhabditis elegans*. Genetics 1993;134:1105–1117. [PubMed: 8375650]

30. Tong AH, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 2001;294:2364–2368.2368 [PubMed: 11743205] A landmark paper that established the high throughput double deletion approach to detecting epistatic interactions.

31. Tong AH, et al. Global mapping of the yeast genetic interaction network. Science 2004;303:808–813. [PubMed: 14764870]

32. Hartman JL, Garvik B, Hartwell L. Principles for the buffering of genetic variation. Science 2001;291:1001–1004. [PubMed: 11232561]

33. Segrè D, Deluna A, Church GM, Kishony R. Modular epistasis in yeast metabolism. Nat Genet 2005;37:77–83. [PubMed: 15592468]

34. Onge RP, et al. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet 2007;39:199–206.206 [PubMed: 17206143] Refs 33 and 34 shows how quantitative information can be incorporated into high throughput interaction studies to yield deeper insights into the nature of genetic networks.

35. Kroll ES, Hyland KM, Hieter P, Li JJ. Establishing genetic interactions by a synthetic dosage lethality phenotype. Genetics 1996;143:95–102. [PubMed: 8722765]

36. Sopko R, et al. Mapping pathways and phenotypes by systematic gene overexpression. Mol Cell 2006;21:319–330. [PubMed: 16455487]

37. Greenspan RJ. The flexible genome. Nat Rev Genet 2001;2:383–387. [PubMed: 11331904]

38. Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. Nat Genet 2006;38:896–903. [PubMed: 16845399]

39. Davierwala AP, et al. The synthetic genetic interaction spectrum of essential genes. Nat Genet 2005;37:1147–1152. [PubMed: 16155567]

40. Tischler J, Lehner B, Fraser AG. Evolutionary plasticity of genetic interaction networks. Nat Genet 2008;40:390–391. [PubMed: 18362882]

41. Wong SL, et al. Combining biological networks to predict genetic interactions. Proc Natl Acad Sci USA 2004;101:15682–15687. [PubMed: 15496468]

42. Beyer A, Bandyopadhyay S, Ideker T. Integrating physical and genetic maps: from genomes to interaction networks. Nat Rev Genet 2007;8:699–710. [PubMed: 17703239]

43. Pattin KA, Moore JH. Exploiting the proteome to improve the genome-wide genetic analysis of epistasis in common human diseases. Hum Genet. 2008

44. Zhu J, et al. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nat Genet 2008;40:854–861.861 [PubMed: 18552845] Shows how interaction information from many sources can be combined to provide a more comprehensive picture of interaction networks.

45. Collins SR, et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature 2007;446:806–810. [PubMed: 17314980]

46. Carlborg O, Jacobsson L, Ahgren P, Siegel P, Andersson L. Epistasis and the release of genetic variation during long-term selection. Nat Genet 2006;38:418–420. [PubMed: 16532011]

47. Stylianou IM, et al. Quantitative trait locus analysis for obesity reveals multiple networks of interacting loci. Mamm Genome 2006;17:22–36. [PubMed: 16416088]

48. Ehrenreich IM, Stafford PA, Purugganan MD. The Genetic architecture of shoot branching in *Arabidopsis thaliana*: a comparative assessment of candidate gene associations vs. quantitative trait locus mapping. Genetics 2007;176:1223–1236. [PubMed: 17435248]

49. Alvarez-Castro JM, Carlborg O. A unified model for functional and statistical epistasis and its application in quantitative trait Loci analysis. Genetics 2007;176:1151–1167. [PubMed: 17409082]

50. Cheverud, JM. Epistasis and the Evolutionary Process. Wolf, J.; Brodie, ED., III; Wade, MJ., editors. Oxford Univ. Press; New York: 2000. p. 58-81.

51. Sambandan D, Yamamoto A, Fanara JJ, Mackay TF, Anholt RR. Dynamic genetic interactions determine odor-guided behavior in Drosophila melanogaster. Genetics 2006;174:1349–1363. [PubMed: 17028343]

52. Causse M, Chaïb J, Lecomte L, Buret M, Hospital F. Both additivity and epistasis control the genetic variation for fruit quality traits in tomato. Theor Appl Genet. 200710.1007/s00122-007-0578-1

53. Rowe HC, Hansen BG, Halkier BA, Kliebenstein DJ. Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. Plant Cell 2008;20:1199–1216. [PubMed: 18515501]

54. Wolf JB, Leamy LJ, Routman EJ, Cheverud JM. Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. Genetics 2005;171:683–694. [PubMed: 16020793]

55. Sinha H, Nicholson BP, Steinmetz LM, McCusker JH. Complex genetic interactions in a quantitative trait locus. PLoS Genet 2006;2:e13. [PubMed: 16462944]

56. Nogami S, Ohya Y, Yvert G. Genetic complexity and quantitative trait loci mapping of yeast morphological traits. PLoS Genet 2007;3:e31. [PubMed: 17319748]

57. Storey JD, Akey JM, Kruglyak L. Multiple locus linkage analysis of genomewide expression in yeast. PLoS Biol 2005;3:e267. [PubMed: 16035920]

58. Brem RB, Storey JD, Whittle J, Kruglyak L. Genetic interactions between polymorphisms that affect gene expression in yeast. Nature 2005;436:701–703.703 [PubMed: 16079846] Shows how genetical genomics can be used to infer patterns of gene interaction.

59. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet 2008;4:e1000008. [PubMed: 18454194]

60. Cheverud JM, Routman EJ. Epistasis and its contribution to genetic variance components. Genetics 1995;139:1455–1461. [PubMed: 7768453]

61. Phillips PC, Johnson NA. The population genetics of synthetic lethals. Genetics 1998;150:449–458. [PubMed: 9725860]

62. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678. [PubMed: 17554300]

63. Tsai CT, et al. Renin-angiotensin system gene polymorphisms and coronary artery disease in a large angiographic cohort: detection of high order gene-gene interaction. Atherosclerosis 2007;195:172–180. [PubMed: 17118372]

64. Wiltshire S, et al. Epistasis between type 2 diabetes susceptibility Loci on chromosomes 1q21-25 and 10q23-26 in northern Europeans. Ann Hum Genet 2006;70:726–737. [PubMed: 17044847]

65. Jamra, R. Abou, et al. The first genomewide interaction and locus-heterogeneity linkage scan in bipolar affective disorder: strong evidence of epistatic effects between loci on chromosomes 2q and 6q. Am J Hum Genet 2007;81:974–986. [PubMed: 17924339]

66. Coutinho AM, et al. Evidence for epistasis between SLC6A4 and ITGB3 in autism etiology and in the determination of platelet serotonin levels. Hum Genet 2007;121:243–256. [PubMed: 17203304]

67. Gregersen JW, et al. Functional epistasis on a common MHC haplotype associated with multiple sclerosis. Nature 2006;443:574–577.577 [PubMed: 17006452] Illustrates how functional hypotheses regarding gene interaction within human populations can be tested using model systems.

68. Trowsdale J. Multiple sclerosis: putting two and two together. Nat Med 2006;12:1119–1121. [PubMed: 17024201]

69. Svejgaard A. The immunogenetics of multiple sclerosis. Immunogenetics 2008;60:275–286. [PubMed: 18461312]

70. Gauderman WJ. Sample size requirements for association studies of gene-gene interaction. Am J Epidemiol 2002;155:478–484. [PubMed: 11867360]

71. Carlson CS, Eberle MA, Kruglyak L, Nickerson DA. Mapping complex disease loci in whole-genome association studies. Nature 2004;429:446–452. [PubMed: 15164069]

72. Xu S, Jia Z. Genomewide analysis of epistatic effects for quantitative traits in barley. Genetics 2007;175:1955–1963. [PubMed: 17277367]

73. Demant P. Cancer susceptibility in the mouse: genetics, biology and implications for human cancer. Nature Rev. Gen 2003;4:721–734.

74. Jacob F. Evolution and tinkering. Science 1977;196:1161–1166. [PubMed: 860134]

75. Lynch M. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc Natl Acad Sci USA 2007;104(Suppl 1):8597–8604. [PubMed: 17494740]

76. Crow JF. How important is detecting interaction? Behavioral and Brain Sciences 1990;13:126–127.

77. Fisher, RA. The Genetical Theory of Natural Selection. Clarendon; Oxford: 1930.

78. Kauffman, SA. The Origins of Order: Self-Organisation and Selection in Evolution. Oxford University Press; New York: 1993.

79. Wu C-I, Palopoli MF. Genetics of postmating reproductive isolation in animal. Annual Review of Genetics 1994;27:283–208.

80. de Visser JA, et al. Perspective: Evolution and detection of genetic robustness. Evolution 2003;57:1959–1972. [PubMed: 14575319]

81. Bridgham JT, Carroll SM, Thornton JW. Evolution of hormone-receptor complexity by molecular exploitation. Science 2006;312:97–101. [PubMed: 16601189]

82. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW. Crystal structure of an ancient protein: evolution by conformational epistasis. Science 2007;317:1544–1548.1548 [PubMed: 17702911] The best example of moving between detailed functional analysis and long term evolutionary inference.

83. Miller SP, Lunzer M, Dean AM. Direct demonstration of an adaptive constraint. Science 2006;314:458–461. [PubMed: 17053145]

84. Weinreich DM, Delaney NF, Depristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 2006;312:111–114. [PubMed: 16601193]

85. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. Nature 2007;445:383–386. [PubMed: 17251971]

86. Karlin S. General two locus selection models: some objectives, results and interpretations. Theoret. Popul. Biol 1975;7:364–398. [PubMed: 1179266]

87. Encode Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007;447:799–816. [PubMed: 17571346]

88. Moore JH, Williams SM. Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. Bioessays 2005;27:637–646. [PubMed: 15892116]

89. Wagner GP, Pavlicev M, Cheverud JM. The road to modularity. Nat Rev Genet 2007;8:921–931. [PubMed: 18007649]

90. Gjuvsland AB, Hayes BJ, Omholt SW, Carlborg O. Statistical epistasis is a generic feature of gene regulatory networks. Genetics 2007;175:411–420. [PubMed: 17028346]

91. Deutscher D, Meilijson I, Kupiec M, Ruppin E. Multiple knockout analysis of genetic robustness in the yeast metabolic network. Nat Genet 2006;38:993–998. [PubMed: 16941010]

92. Jansen RC. Studying complex biological systems using multifactorial perturbation. Nat Rev Genet 2003;4:145–151.151 [PubMed: 12560811] A nice perspective on how complex genetic systems can be best interrogated using multiple, rather than single, perturbations.

93. Carter GW, et al. Prediction of phenotype and gene expression for combinations of mutations. Mol Syst Biol 2007;3:96. [PubMed: 17389876]

94. Bateson, W. Mendel's Principles of Heredity. Cambridge Univ. Press; Cambridge: 1909.

95. Fisher RA. The correlations between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinb 1918;52:399–433.

96. Tachida H, Cockerham CC. A building block model for quantitative genetics. Genetics 1989;121:839–844.844 [PubMed: 2721934] A greatly underappreciated paper that provides a quantitative framework for moving between different perspectives for how phenotypes are built and how genetic effects can be estimated.

97. Karlin S, Feldman MW. Simultaneous stability of D=0 and D≠0 for multiplicative viabilities at two loci. Genetics 1978;90:813–825. [PubMed: 17248876]

98. Mani R, Onge RP, Hartman J. L. t. Giaever G, Roth FP. Defining genetic interaction. Proc Natl Acad Sci U S A 2008;105:3461–3466.3466 [PubMed: 18305163] Shows how dependent the inference of epistasis is upon the scale of measurement.

99. Aylor DL, Zeng ZB. From classical genetics to quantitative genetics to systems biology: modeling epistasis. PLoS Genet 2008;4:e1000029. [PubMed: 18369448]

100. Feldman MW, Otto SP, Christiansen FB. Population genetic perspectives on the evolution of recombination. Annual Review of Genetics 1997;30:261–295.

101. Bennett DC, Lamoreux ML. The color loci of mice-a genetic century. Pigment Cell Res 2003;16:333–344. [PubMed: 12859616]

102. Steiner CC, Weber JN, Hoekstra HE. Adaptive variation in beach mice produced by two interacting pigmentation genes. PLoS Biol 2007;5:e219. [PubMed: 17696646]

103. Silvers, W. The Coat Colars of Mice. Springer-Verlag; Berlin: 1979.

104. Hoekstra HE. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity 2006;97:222–234. [PubMed: 16823403]

105. Campbell, NA.; Reece, JB.; Taylor, MR.; Simon, EJ.; Dickey, JL. Biology: Concepts and Connections. Benjamin Cummings; New York: 2008.

106. Wright S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc. 6th Int. Cong. Genet 1932;1:356–366.

107. Gavrilets, S. Fitness landscapes and the Origin of Species. Princeton University Press; Princeton, NJ: 2004.
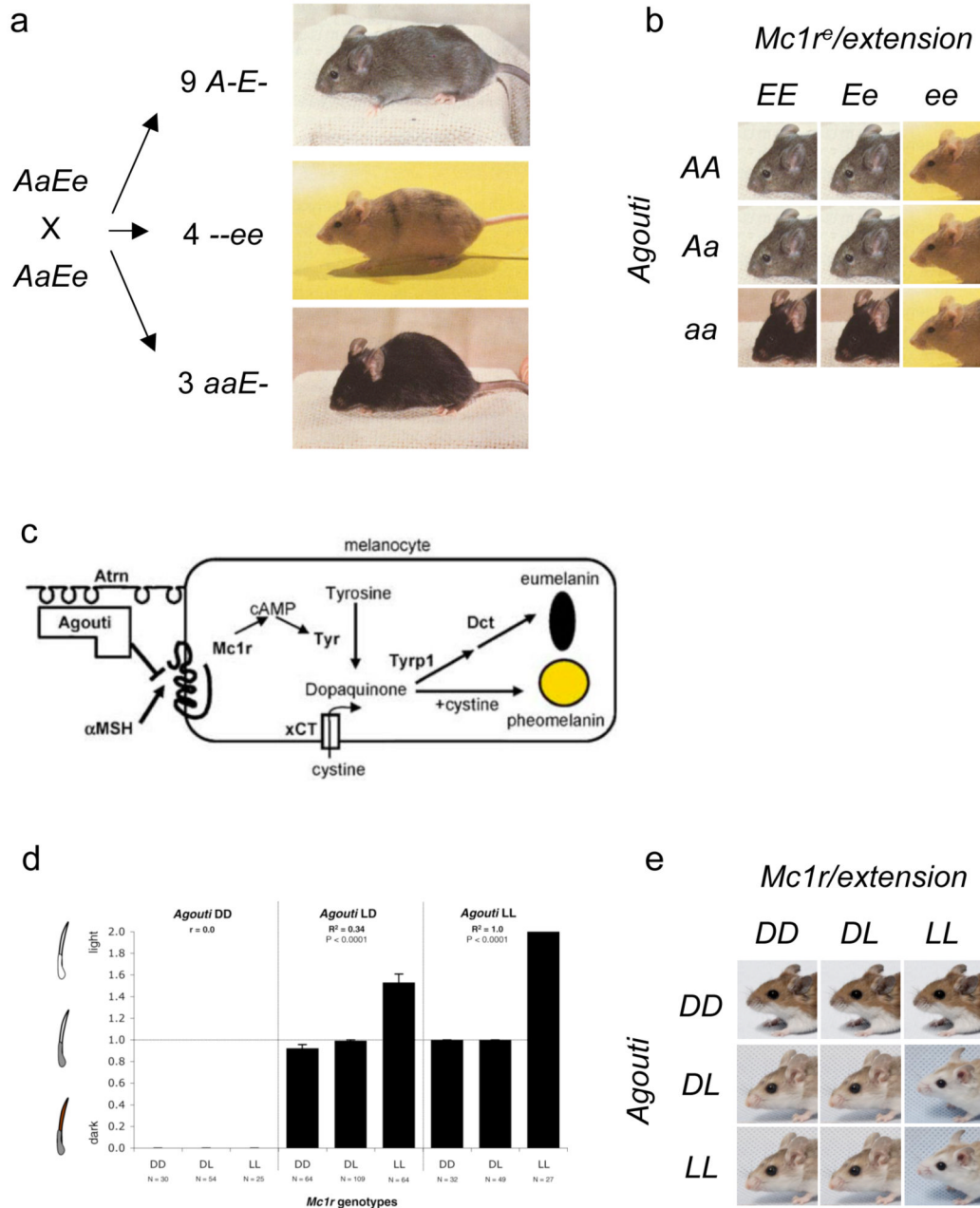
Figure 1.
Different viewpoints of epistasis. Coat color variation in mammals has long served as one of most fruitful examples of genetics, with over 120 loci and 800 alleles described in mice alone[101]. The coat color genetic pathway can be used to illustrate different usages of the term "epistasis". In the original sense, from Bateson, epistasis arises when the effects of alleles at one locus are blocked by the presence of a specific allele at another locus. (**a**) For example, a cross between an agouti and extension (now called the melancortin 1 receptor or *Mc1r*) double heterozygotes yields the non-Mendelian segregation ratio of 9:4:3 (instead of 9:3:3:1), with the excess of dusty yellow extension offspring suggesting that the *Mc1r* locus operates downstream of agouti (an example of recessive epistasis). Crosses with other mutants can be

used to further order the genetic pathway, relying on combinations of knockout mutants to generate *compositional epistasis* (Box 1). (**b**) The outcomes of this same cross can be illustrated in a 3×3 genotype interaction table that is common in population genetics. (**c**) The biochemistry of this pathway within the melanocyte has since been fully worked out. Here, activation of *Mc1r* turns on the production of the (dark) eumelanin, as opposed to (the yellow/red) pheomelanin. *Agouti* acts an antagonist to *Mc1r* such that periodic activation of *agouti* leads to banding of color on individual hairs. Disruption of other loci, such as the classic *albino* (*Try*) locus, destroys function of the entire pathway. The specific interactions between the proteins in this pathway are representative of *functional epistasis*. (**d**) Recently, Steiner et al. [102] have used this pathway to probe natural variation between dark forest mice and light colored beach mice. They find that the adaptive transition from dark to light that accompanied the movement of mice from the forest to the beach is accomplished by an interaction between structural changes to the agouti locus and regulatory changes to the *Mc1r* locus. The specific effects of these interactions can be roughly quantified by averaging the effects of these markers overall all sampled genetic backgrounds to give an estimate of the *statistical epistasis* between these loci. (**e**) Interestingly, the pattern of epistasis for these loci in nature is reversed from the standard cross, presumably because the *Mc1r* allele in the beach mice has partial function and is therefore still susceptible to suppression from agouti. This later observation is a clear illustration that epistasis is a property of specific alleles, rather than a particular locus in general. Pictures for (**a**, **b**) reprinted with permission from Ref 103. Images are representative of similar phenotypes. (**c**) is modified from Ref 104). (**d**) is from Ref 102. Images for (**e**) kindly supplied by Hopi Hoekstra.
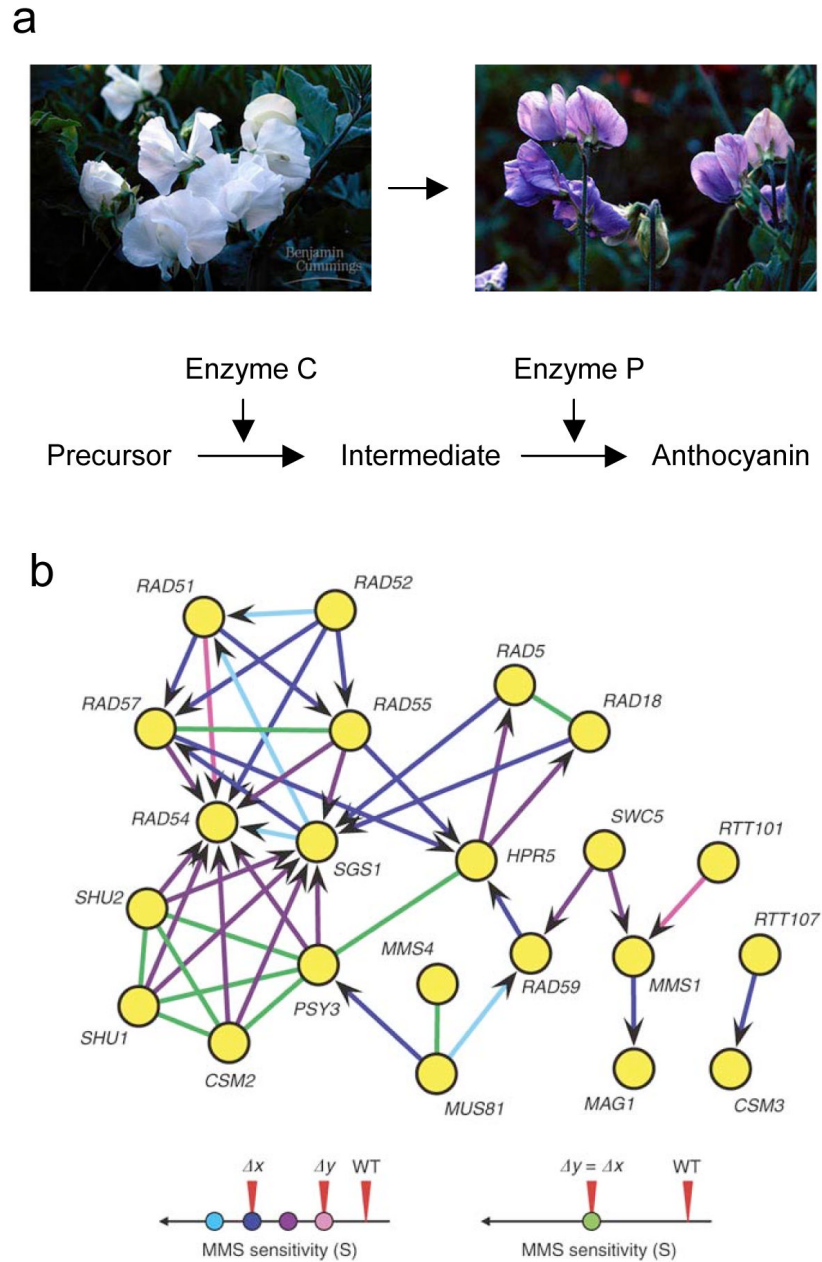
Figure 2.
Reconstructing genetic pathways using epistasis analysis. (**a**) A classical example of epistasis using flower color in sweet peas. Combinations of mutations at two loci generating enzymes responsible for processing anthocyanin operate within a single biochemical pathway. Picture reprinted from Ref 105. (**b**) Construction of the epistatic network of genes underlying sensitivity to the mutagen methyl methanesulfanate (MMS) in yeast. The fitness of all 650 combinations of 26 genes known to be sensitive to MMS were measured quantitatively; combinations showing departures from multiplicative interactions are connected in the network. Arrows represent five interaction subtypes: coequal (green), partial masking (light

blue), masking (dark blue), partial suppression (purple), suppression (pink). Reprinted from Ref 34.
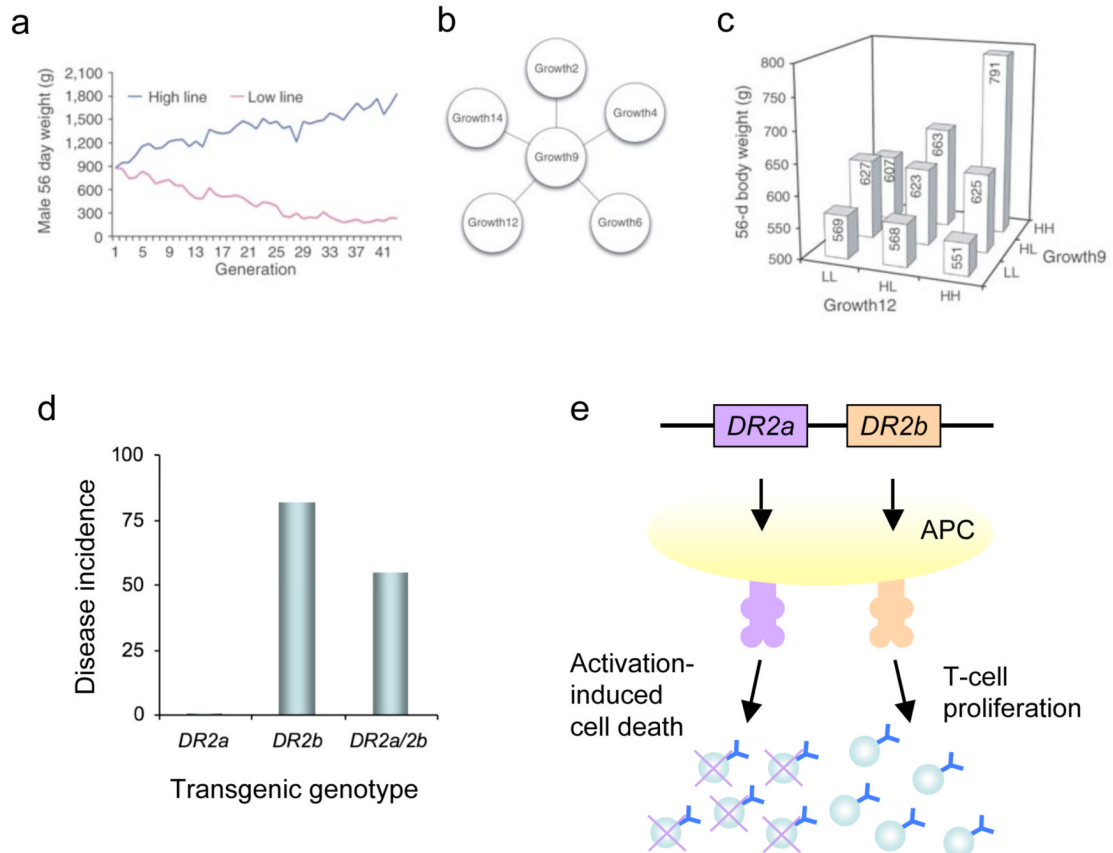
Figure 3.
Epistasis in complex traits. (**a**) More than forty generations of selection on body size in chickens lead to a large difference among lines. (**b**) QTL mapping of an F2 cross between the divergent body size lines revealed a network of six genes that explains 45% of the among line difference. Five of these loci are only show their effects via epistasis with the sixth (Growth9). (**c**) An example of one of the genotype/phenotype relationship between two of these genes. Note the large difference in the double Growth12/Growth9 high line homozygotes. (**d**) Epistasis in the autoimmune reaction thought to underlie many cases of multiple sclerosis. Mice that had been transformed to produce human T-cell antigen receptors, are then also subsequently transformed to express the human major histocompatibility (MHC) locus DRB5*0101 (*DR2a*), the MHC locus DRB1*1501 (*DR2b*), or both. The incidence of disease is much lower when both loci are present, even though the *DR2a* locus does not independently influence disease onset. (**e**) These two loci encode antigen binding proteins that are expressed on the surface of an antigen presenting cell (APC). A possible model for the epistatic effect is that the overall antigenic effect is maintained by a balance between T-cell proliferation induced by *DR2b* and T-cell apoptosis induced by *DR2a*. The complex relationship between these two might explain the complex pattern of onset and periodic nature of some multiple sclerosis cases. (**a-c**) from Ref 46. (**d**) is based on data from Ref 67. (**e**) is adapted from Ref 68.
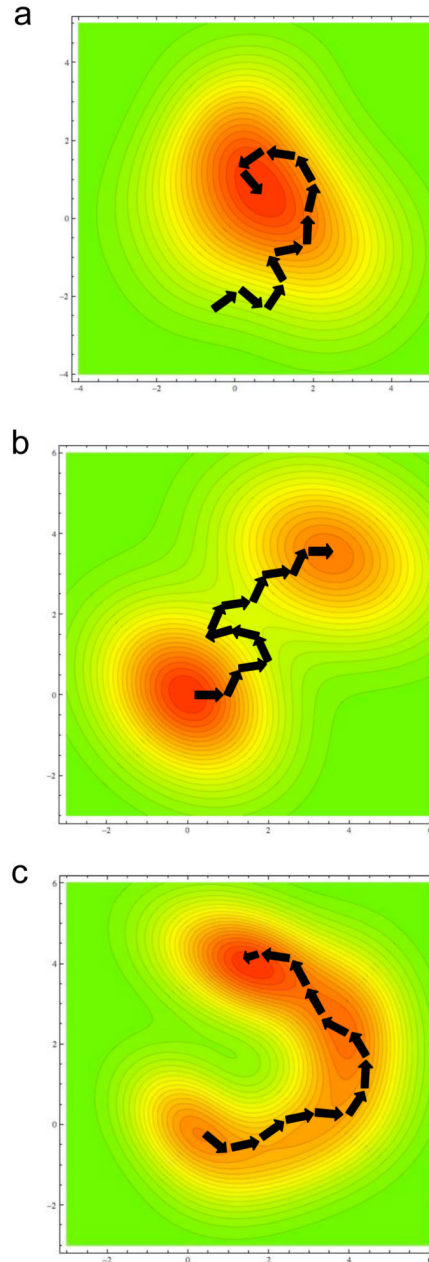
Figure 4.
Four different views of the generation of epistasis under natural selection. Each figure shows a hypothetical adaptive landscape in which the mean fitness of a population is a function of underlying variation at two or more loci. Lines display points of equal fitness, with warmer colors indicating areas of higher fitness. (**a**) Under **Fisherian Adaptation**[77], adaptive evolution is seen as a hill climbing process in which new beneficial mutations that increase fitness build upon substitutions that are already fixed within the population. Here, epistasis is a byproduct of the historical contingency built into the evolutionary process (both from the random effects of mutations and shifts in the adaptive optimum), but at any given time point the average or additive effects of a particular mutation are responsible for the actual

evolutionary change. (**b**) In **Wright's Shifting Balance Process**[106], epistasis generates multiple adaptive peaks in the fitness landscape. If stable through time, these valleys can only be traversed via genetic drift or tight linkage. Here, epistasis plays a fundamental role in determining the direction and likelihood of any particular evolutionary change. (**c**) In **Gavrelet's Holy Landscape** model[107], adaptive landscapes are rugged as in Wright's case, but the multidimensional nature of the fitness landscape leads to ridges of nearly equal fitness that populations move along either via directional selection or genetic drift. This represented metaphorically here by a ridge of nearly equal fitness whose endpoints are spatially separated by a deep valley. In multidimensional space, this might be something more like a tunnel connecting regions of high fitness and avoiding complex adaptive valleys. Here, epistasis will be strongly apparent after populations diverge, and may impact the direction of evolutionary change, but does not generate a barrier to change as in Wright's model.