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

Patrick C. Phillips

Abstract | Epistasis, or interactions between genes, has long been recognized as fundamentally important to understanding 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 new-found 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 versus entire systems — this is an old topic of debate within genetics, but 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 was 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 part that gene interactions play within biological systems.

It has been approximately 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 use of the term 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, the genetic ordering of regulatory pathways and the quantitative differences of allele-specific effects (FIG. 1). Using the same word to describe subtly different phenomena has generated surprisingly little confusion in the literature — mostly because of the tendency for different areas of genetics to almost completely ignore one another. This is no longer possible. Molecular geneticists are now studying how specific allelic effects traverse complex regulatory networks, and evolutionary geneticists are moving from statistical descriptions of genetic variation to identification of the specific nucleotide changes that are 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, often in complex ways. The challenges that are generated by the presence of epistasis in a system provide a focal point for the unification of traditionally disparate areas of research and show that this fundamental genetic concept is more relevant now than it ever has been.

In this article I first review various definitions of epistasis and show what they have in common and how they differ. I then turn to how the analysis of epistatic interactions between genes can be used to elucidate the global structure of these systems. I will also examine the impact that epistasis has on our ability to understand the genetic basis of natural variation, especially how it pertains to genetic variation associated with disease within human populations. Finally, I review various models of how evolution builds complex systems and I explore how recent studies of molecular evolution can be used to determine the role that epistasis has in directing the path of evolutionary change. As this Review has such a broad scope, it will not be possible to cover

Center for Ecology and Evolution, University of Oregon, Eugene, Oregon 97403 USA, and Instituto Gulbenkian de Ciência, Apartado 14, P-2781-901 Oeiras, Portugal.
e-mail: pphil@uoregon.edu
doi:10.1038/nrg2452
Published online
14 October 2008

© 2008 Macmillan Publishers Limited. All rights reserved

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 observed and 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 were called 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.

Perhaps it was natural, therefore, 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 (for example, during segregation in a specific cross) does not necessarily mean that there will be epistasis in the other context (for example, 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 complete system[96]? The first approach 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, as the mutant and wild-type 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? Because there is no such thing as a 'naked' gene with no 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).

all of the relevant literature. However, there has 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 led to a plethora of differently nuanced meanings for the term epistasis, all of which involve gene interactions at various levels (BOX 1). Although a few scientists have suggested using the more generic term 'gene interaction' to encompass the variety of phenomena labelled as epistasis, so that epistasis can retain its original, more specialized meaning[1], this seems untenable given the history of use. As is evident below, traditional uses of epistasis to order genes within pathways have become increasingly quantitative, further obscuring the boundaries of its definition and suggesting that expansion rather than contraction of the usage of epistasis is required. The uses of the term epistasis can be condensed into three main categories, each of which have been defined in many ways, so I impose my own labels (BOX 1; FIG. 1).

'Functional epistasis' addresses the molecular interactions that proteins (and other genetic elements) have with one another, whether these interactions consist of proteins that operate within the same pathway or of proteins that 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 between the proteins were to be disrupted. Because the use of the term epistasis is confusing enough 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.

'Compositional epistasis' is a new term that is 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 manifested, 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 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 that are exposed in the double mutant homozygote, then we essentially have the definition of epistasis that is being used in modern systems biology. This does not have to be limited to qualitative categories of phenotypes: it can also include quantitative measures of interaction against a fixed background (BOX 2).

'Statistical epistasis' is the usage of epistasis that is attributed 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 manifested, so there is no such thing as a standard genetic background within a natural population, only the set of backgrounds that is 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

© 2008 Macmillan Publishers Limited. All rights reserved

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 that they are sampled from a population as opposed to being intentionally constructed, as is described by compositional epistasis.

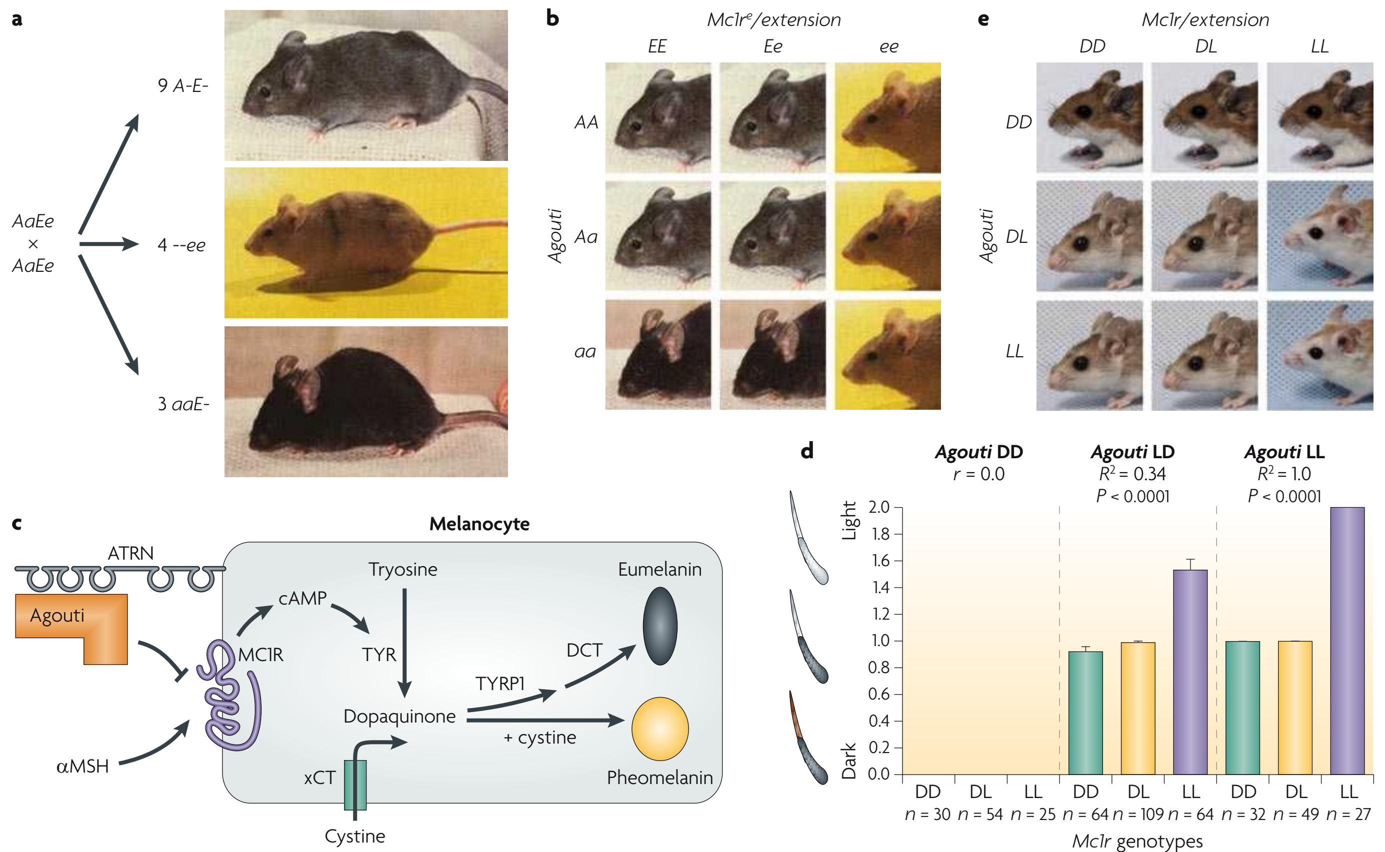


Figure 1 | **Different viewpoints of epistasis.** Coat colour variation in mammals has long been is one of the most fruitful examples in the study of the relationship between genotype and phenotype, with over 120 loci and 800 alleles described in mice alone[101]. The coat colour genetic pathway can be used to illustrate different usages of the term epistasis. In the original sense, defined by Bateson, epistasis arises when the effects of alleles at one locus are blocked by the presence of a specific allele at another locus. For example, a cross between agouti and extension (now called the melanocortin 1 receptor or *Mc1r*) double heterozygotes (AaEa) yields the non-Mendelian segregation ratio of 9:4:3 (instead of 9:3:3:1) (**a**), with the excess extension offspring (highlighted by a yellow background) suggesting that the *Mc1r* locus operates downstream of *agouti* (an example of recessive epistasis). Crosses with other mutants can be used to order other components of the genetic pathway, relying on combinations of knockout mutants to generate compositional epistasis (BOX 1). The outcomes of this same cross can be illustrated in a 3x3 genotype interaction table that is common in population genetics, images are representative of similar phenotypes (**b**). The biochemistry of this pathway within the melanocyte has since been fully elucidated (**c**). Activation of MC1R turns on production of eumelanin, as opposed to the default production of pheomelanin. The agouti protein acts an antagonist to MC1R, leading to periodic activation of the agouti protein and banding of colour on individual hairs. Disruption of other loci, such as the classic tyrosinase locus (*Tyr*; previously known as albino), destroys function of the entire pathway. The specific interactions between the proteins in this pathway are representative of functional epistasis. Recently, Steiner *et al.*[102] have used this pathway to probe natural variation between dark-coloured forest mice and light-coloured beach mice (**d**). They found that the adaptive transition from dark (D) to light (L) 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 approximately quantified by averaging the effects of these markers over all sampled genetic backgrounds to give an estimate of the statistical epistasis between these loci. Interestingly, the pattern of epistasis for these loci in nature is reversed from the standard cross (**e**), presumably because the *Mc1r* allele in the beach mice has partial function and is therefore still susceptible to suppression from *agouti*. This observation is a clear illustration that epistasis is a property of specific alleles, rather than a particular locus in general. αMSH, α-melanocyte-stimulating hormone; ATRN, attractin; DCT, dopachrome tautomerase; TYRP1, tyrosinase-related protein 1; xCT, solute carrier family 7 (cationic amino acid transporter, y+ system), member 11. Images in parts **a**, **b** are reproduced, with permission, from REF. 103 © (1979) Springer-Verlag. Images in part **c** are modified, with permission, from REF. 104 © (2006) The Genetics Society. Part **d** is modified from REF. 102. Part **e** is courtesy of Hopi Hoekstra, Harvard University, USA.

© 2008 Macmillan Publishers Limited. All rights reserved

Box 2 | Measures of epistasis

In principle, detecting epistasis using Bateson's definition (BOX 1) is 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 more complex. First of all, epistasis means that something different happens when a particular set of alleles from different loci are found in combination than when they are 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 could be represented as $W_{xy} = \alpha_x + \alpha_y + \varepsilon$, where α is the individual effects of each allele at loci x and y, ε is the deviation that is due to epistasis and W is the observed phenotype. Relationships for diploids are more complex because of the possibility of one locus interacting with the dominance state of the other locus[14]. Fisher presumably chose this definition because additive linear models are 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 1960s, population geneticists started using deviation from a multiplicative model of gene action as the definition of epistasis, instead of deviation from 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 be represented as $W_{xy} = \alpha_x \alpha_y + \varepsilon$. It is an underappreciated 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 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.

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 this expectation depends on how one measures epistasis. The stability of linkage disequilibrium depends strongly on the recombination rate, especially when linkage is tight[97]. Now that we have a firm idea of the genomic structure of many organisms, we know that there can often be 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, there is a range 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 phenomena, which makes it difficult to understand what individual researchers mean when they use these terms. For example, synergistic epistasis 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 deleterious mutations then the phenotype is less than expected, but for positive mutations the phenotype is greater than expected. So sometimes synergistic epistasis means 'extra good' and sometimes it means 'extra bad'. The field would benefit if all of these terms, which have context-dependent meanings, were replaced with two simple terms: positive epistasis 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 two terms are preferable because their meaning is immediately clear and because it is the sign of the epistasis that matters in most evolutionary processes (such as the generation of linkage disequilibrium), not the change in relative direction of the effects of the individual loci[2]. This change in relative direction can be addressed using another simple term, sign epistasis[5]. This term indicates that the direction of the epistatic and individual effects differ from one another and that the direction of selection on the individual alleles changes depending on the genetic context. For example, if two mutations lower fitness when found individually but increase fitness when found together, then this results in an adaptive valley, which has different functional and evolutionary implications than if the sign of the individual effects and epistasis are in the same direction (FIG. 4).

Admixture
The pattern of genetic variation that occurs when a population is derived from founders that originated from more than one ancestral population.

Punnett square
A method of calculating the outcomes of a genetic cross by multiplying the expected frequency of gametes from a mother by the expected frequency of gametes from the father.

Hardy–Weinberg equilibrium
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 disequilibrium) the genotype frequencies are equal to the product of the allele frequencies.

The last two views of epistasis are complementary to one another. Compositional epistasis measures the effects of allele substitution against a particular fixed genetic background, whereas 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 might make them a transient statistic. However, Fisher would undoubtedly point out that any measure of interaction 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; therefore, compositional epistasis might be considered an arbitrary measurement. There is a compromise between precision and generality when comparing the two types of epistasis. Statistical measures are often an average from variable epistatic effects at many different loci, which might lead to these effects being cancelled out[2,21].

The hierarchical structure of the relationship between compositional and statistical epistasis is similar, for example, to the way that a Punnett square might be seen as a special case of the Hardy–Weinberg equilibrium 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 favour these terms in this context because they are still essentially statistical measures and do not tell us

© 2008 Macmillan Publishers Limited. All rights reserved

anything about gene function in the way that, for instance, a molecular biologist would use the term.

Ultimately, different uses of epistasis can be unified under a single perspective, using the view that epistasis measures biallelic substitution under different genetic backgrounds, whether they are fixed or average. It is as simple — and as complex — as that.

Epistasis as a tool

One of the first characteristics studied by Bateson and Punnett[23] that revealed a pattern of epistasis was flower colour in sweet peas. As illustrated in nearly every genetics textbook, it is possible to cross two colourless (white) flowers and recover purple flowers in the offspring. The non-Mendelian segregation ratios of the F2 plants 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 in the anthocyanin pathway, such that a mutation in either gene can disrupt flower colour.

In general, the fact that the phenotype of an individual depends strongly on the specific combination of alleles at two or more loci suggests that this dependency must be informative 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 *Caenorhabditis 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 larva 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 be used to make predictions about the probable functions of the identified genes and be tested by the molecular characterization of the gene products.

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 they result in death when both are present, whereas an individual with either mutation alone is viable.

High-throughput approaches. The vast majority of studies that have used epistasis to analyse 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 pairwise interactions between all genes? The problem here, of course, is one of scale. The number of pairwise interactions between genes grows at approximately the square of the number of genes: $n(n - 1)$ to be exact, where n is the number of loci, or $n(n + 1)$ if the parental strains are included. These numbers are halved if reciprocal interactions do not need to be tested. This type of study is a daunting but achievable task for the 190 interactions resulting from 20 genes, but would be close to 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 knockdown libraries, and of high-throughput maintenance and screening techniques.

One among the first studies of this type was performed by Tong *et al.*[30], who used a high-throughput analysis approach (synthetic genetic array analysis) to examine the interactions of 8 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 in this study define a network of ~1,000 genes involving ~4,000 interactions, with most interactions tightly grouped in self-similar functional clusters. These results help us understand how complex gene networks can build robustness into cellular systems[32] but, because they are based on a growth versus no growth criterion, they are largely qualitative[33]. St. Onge *et al.* provide a good example of how classical compositional epistasis can be combined with a more quantitative approach in their examination of the genetic interaction system that influences the resistance of yeast to the mutagen methyl methanesulphonate (MMS)[34]. Using 26 mutants that are known to influence MMS resistance, St. Onge *et al.* constructed all 650 possible double-deletion strains. Of these, 10 interactions generated synthetic-lethal mutations, 67 were classified as 'aggravating interactions' (negative epistasis; BOX 2) and 45 were classified as 'alleviating interactions' (positive epistasis), on the basis of a multiplicative model. The interactions within the positive epistasis class, which provide greater than expected resistance to MMS, were used to generate a functional interaction map among the loci (FIG. 2). Most of the already known genetic pathways were identified, 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 (that is, a coequal relationship) involved direct interactions between protein subunits, suggesting that it might be possible to connect particular epistatic 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, which is certainly an extreme approach to perturbing complex systems. For example, Kroll *et al.*[35] devised a method for looking for interactions that are induced after systematically overexpressing genes. Using this approach, Sopko *et al.*[36] found that, when overexpressed in *Saccharomyces cerevisiae*, about 15% of a set of 5,280 yeast genes induced a growth defect, with most of the overexpression effects not matching the phenotypes of their corresponding deletions. Testing a deletion of the cyclin-dependent kinase gene *pho85* across the overexpression library revealed 65 synthetic interactions,

© 2008 Macmillan Publishers Limited. All rights reserved

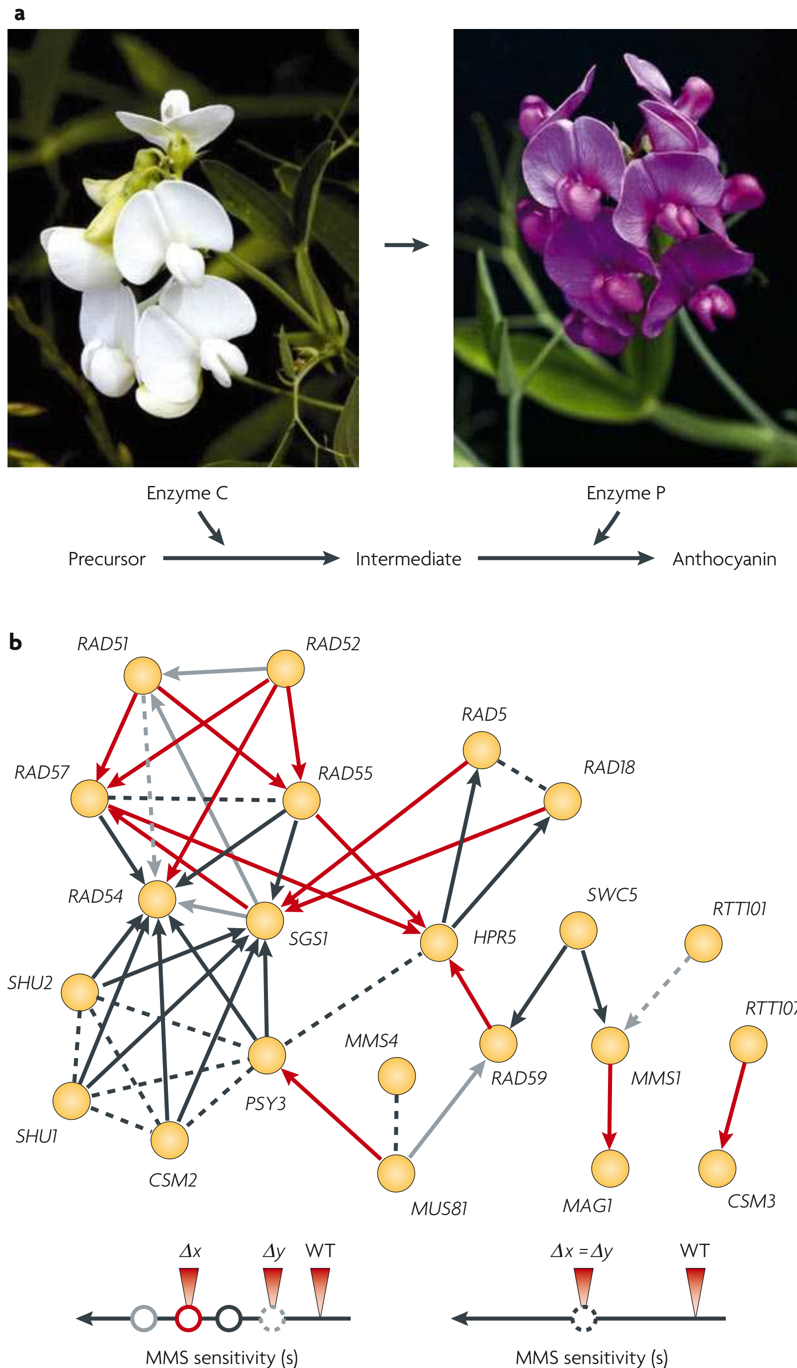


Figure 2 | **Reconstructing genetic pathways using epistasis analysis. a** | A classical example of epistasis using flower colour in sweet peas. Combinations of mutations at two loci that encode enzymes that are responsible for processing anthocyanin operate within a single biochemical pathway. **b** | Construction of the epistatic network of genes underlying sensitivity to the mutagen methyl methanesulphonate (MMS) in *Saccharomyces cerevisiae*. 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 (dashed black lines), partial masking (solid grey lines), masking (solid red lines), partial suppression (solid black lines) and suppression (dashed grey lines). WT, wild type; Δx, effect of deletion at first locus; Δy, effect of deletion at second locus. Photos in part **a** are reproduced, with permission, from http://www.photographybydave-lines.com © Dave Lines. Part **b** is modified, with permission, from REF. 34 © (2007) Macmillan Publishers Ltd. All rights reserved.

most of which were previously unknown. There are of course an endless combination of knockout, overexpression, natural and induced alleles that can be combined to probe a given library or array. The fact that the specific interaction results obtained depend on the nature of the probe (deletion, overexpression and so on) 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', in which perturbation of 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 *S. cerevisiae* a particularly powerful system for systematically dissecting epistatic interaction networks (reviewed in REFS 18,19). Unfortunately, such approaches are 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 knock down (rather than knock out) genes in a systematic fashion. Lehner *et al.*[38] used 37 strains of *C. elegans* with mutations in cell-signalling components to query an RNAi library of ~1,750 genes involved in signal transduction, transcriptional regulation and chromatin remodelling. The query was executed by raising each strain on bacteria that express 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. Although the number of interactions might seem low considering the total number that are possible, it is on the same order as that observed in *S. cerevisiae*, 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 seems that the structure of the interaction networks between yeast and worms is somewhat 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 currently fruitful approach seems to be to concentrate on a large but finite set of genes that are known to be involved in a well-defined biological process. For example, Collins *et al.*[45] used all of the pairwise interactions of the 743 genes known to influence chromosomal processes such as DNA repair, transcriptional regulation and

© 2008 Macmillan Publishers Limited. All rights reserved

chromatid segregation in yeast. This allowed them to place their interaction results in the context of already well-explored systems, such as the biochemistry of the multiprotein Mediator transcriptional co-activation complex. In this case, epistatic interactions corresponded well to known physical interactions among proteins, but also 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 finding ways to overcome the inherent scaling problem of the exponential growth in the number of possible genetic interactions. Furthermore, now that we are beginning to look at pairwise interactions, how do we know whether third- or higher-order interactions will 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. 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, the use of which is sure to continue to grow.

Epistasis as an obstacle

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 QTL mapping, is to start with a given set of phenotypes and then to identify the genes that are 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 illustrated by a recent study by Carlborg *et al.*[46] into the genetic basis of the response to selection on body size in chickens (FIG. 3). After approximately 40 generations of selection, males in the low-weight line weighed six times less than males in the high-weight line. It would seem likely that there would be a strong genetic signal behind such a difference. However, after examining many marker loci for their individual effects, only one QTL (named Growth9) seemed to have an effect, and the signal for that was 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 in the high-growth background, that is, the line with the Growth9 QTL (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 completely accounted for by its epistatic interactions with the other QTLs. A similar pattern of modules of interacting QTLs has been identified as influencing obesity in mice[47]. Although the genes underlying these QTLs still need to be identified, it is clear that the vast majority of the genetic information in this system would have been missed if this study had not also looked 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 might 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 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 odour-sensing behaviour in *Drosophila melanogaster*[51], growth and yield in tomatoes[52], the *A. thaliana* 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 the residual variation segregating within natural populations that determines differences among individuals, not the total scaffold of biological function. 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 has 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 dissect the specific genetic basis of complex traits, will epistasis have a larger part to play? The answer is that we still do not know. This is partly 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 partly because the statistical issues involved in estimating large numbers of potential interaction effects has limited the power of most existing studies. However, 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. So it is 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.

Chromatin immunoprecipitation
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.

© 2008 Macmillan Publishers Limited. All rights reserved

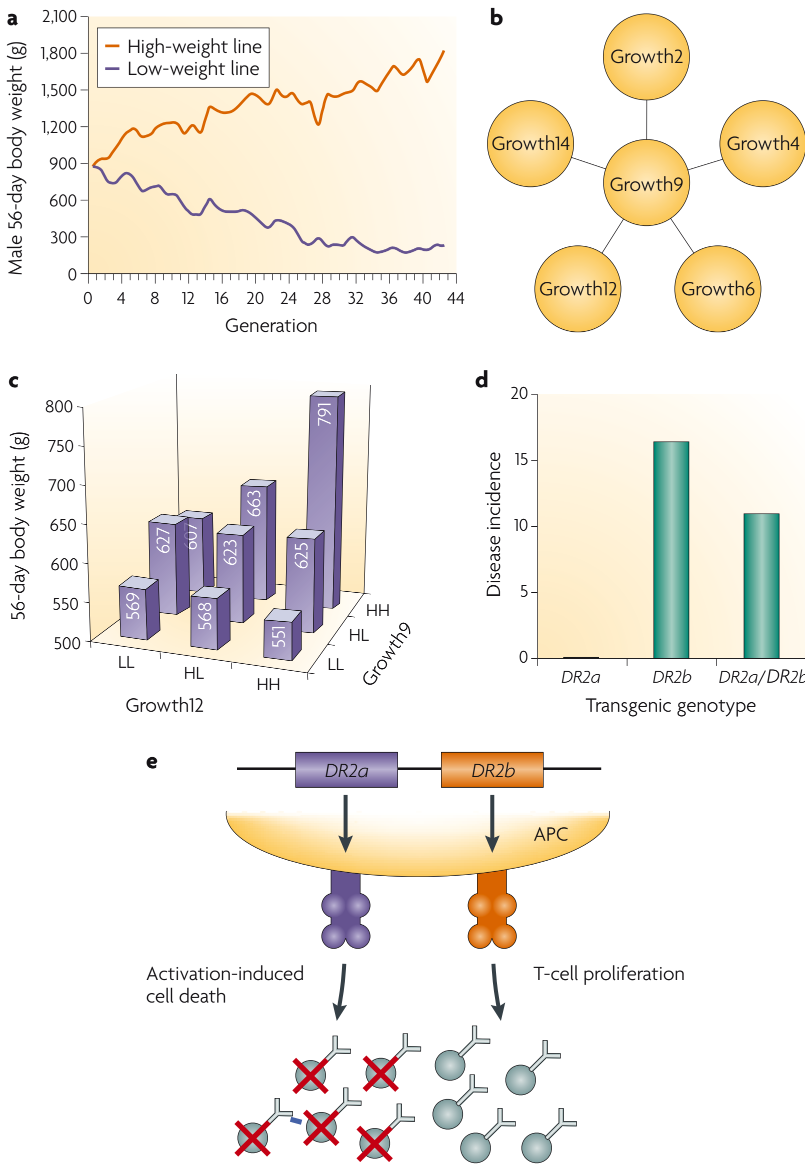


Figure 3 | **Epistasis in complex traits. a** | More than 40 generations of selection on body weight in chickens has led to a large difference among lines. **b** | QTL mapping of an F2 cross between the divergent body-weight lines revealed a network of six genes that explains 45% of the among-line difference. Five of these loci only show their effects through epistasis with the sixth (Growth9). **c** | An example of one of the genotype–phenotype relationships between two of these genes. Note the large difference in the double Growth12–Growth9 high-weight line homozygotes (HH). **d** | Epistasis in the autoimmune reaction that is 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) loci *HLA-DRB5*0101* (*DR2a*), *HLA-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 loci might explain the complex pattern of onset and the periodic nature of some multiple sclerosis cases. Parts **a**–**c** are modified, with permission, from REF. 46 © (2006) Macmillan Publishers Ltd. All rights reserved. Part **d** is based on data from REF. 67 © (2006) Macmillan Publishers Ltd. All rights reserved. Part e is modified, with permission, from REF. 68 © (2006) Macmillan Publishers Ltd. All rights reserved.

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 question of whether previous problems are a function of limitations in the data or are 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 of these cases involves the genetic interactions underlying the autoimmune disease multiple sclerosis. Here, Gregersen *et al.*[67] found evidence that natural selection might be maintaining linkage disequilibrium between the histocompatibility loci *HLA-DRB5*0101* (*DR2a*) and *HLA-DRB1*1501* (*DR2b*) (FIG. 3), which are known to be associated with multiple sclerosis; linkage disequilibrium can be generated by strong epistasis among adjacent loci (BOX 2). To test the idea that epistasis is occurring between *DR2a* and *DR2b*, Gregersen *et al.* generated genetically engineered mice that synthesize the corresponding human immune proteins and found that mice producing the protein product of *DR2b* were highly susceptible to disease, whereas those producing the *DR2a* product did not progress towards disease. Then, in a crucial test, mice expressing both alleles 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 that are sensitive to the antigen that induces multiple sclerosis, 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: under most conditions the influence of these two factors balance 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. This is because the natural recombinants between *DR2a* and *DR2b* that would be required for such an analysis have yet to be observed[68]. These issues only serve to highlight how difficult it can be to identify underlying complex diseases, even without 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 that would be required suggests that stringent significance thresholds will be needed to

© 2008 Macmillan Publishers Limited. All rights reserved

Linkage disequilibrium
A measure of whether alleles at two loci coexist in a population in a non-random fashion. Alleles that are in linkage disequilibrium are found together on the same haplotype more often than would be expected under a random combination of alleles.

control against false positives, but this in turn means that the only epistatic effects that will be detected will have to be huge and/or sample sizes will need to be very large[70]. Indeed, larger studies are more likely to detect epistasis than smaller ones[6]. Any suggestions to limit testing to only those QTLs with significant main effects[71] are probably ill-advised: epistatic interactions with the largest relative effect sizes will be those with small main effects and we already know that epistasis is frequently detected in the absence of main effects[50,72]. Carlborg and Haley[6] advocate a sequential approach in which potentially interesting QTLs are first identified using a high false-discovery rate, and are then used for subsequent tests for genetic interactions. Specialized breeding designs can be used to increase the resolution and the 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[42,43]. Although there is 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 a limited phenomenon.

Structure and evolution of complex systems

Where does all of this epistasis come from? 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 an extremely likely outcome of the evolutionary process. Because 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 when new effector molecules attach themselves to the backbone of an existing signal transduction pathway, for example. 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, which is epistatic to many other genes, can bring the whole structure down. This is 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

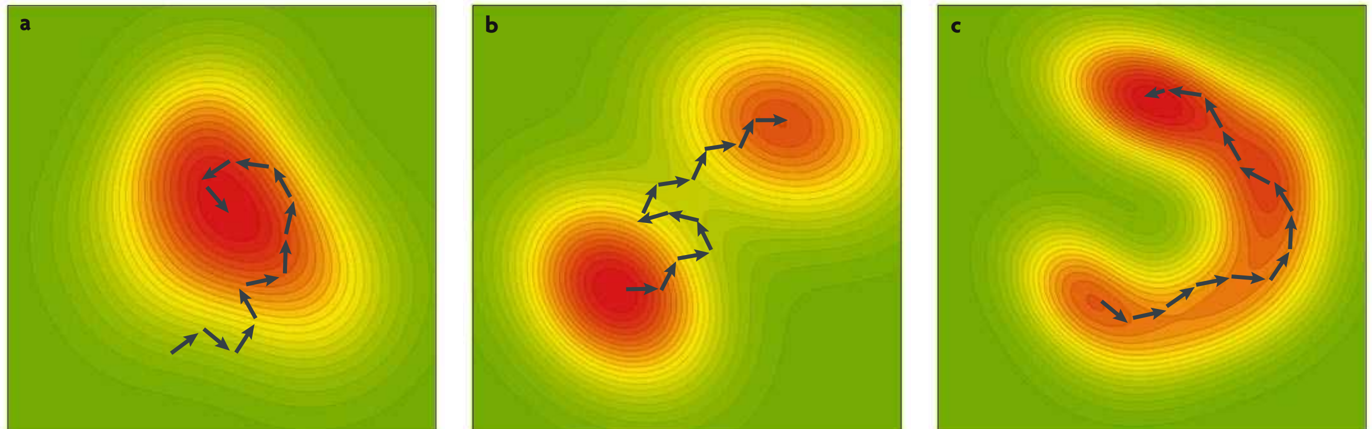
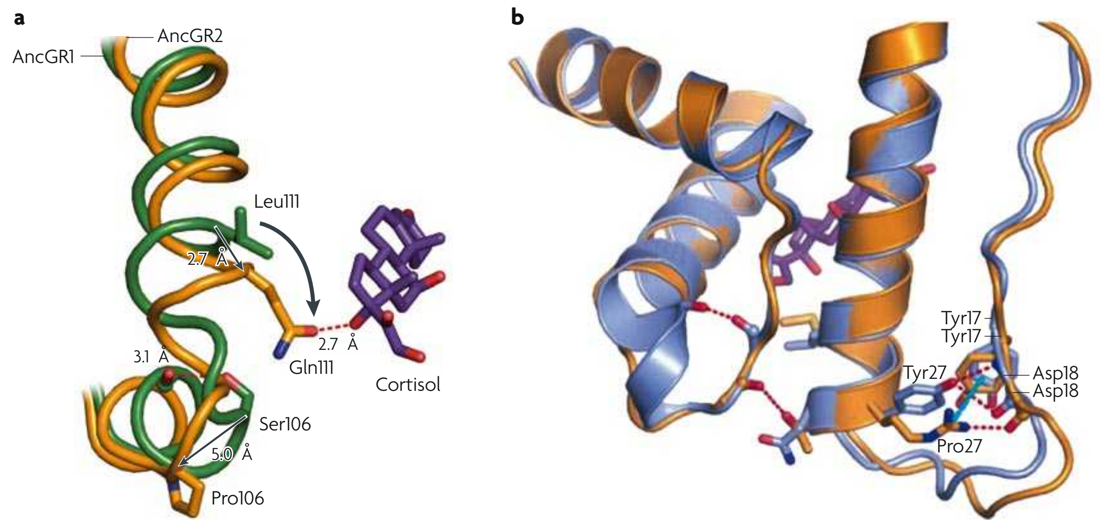
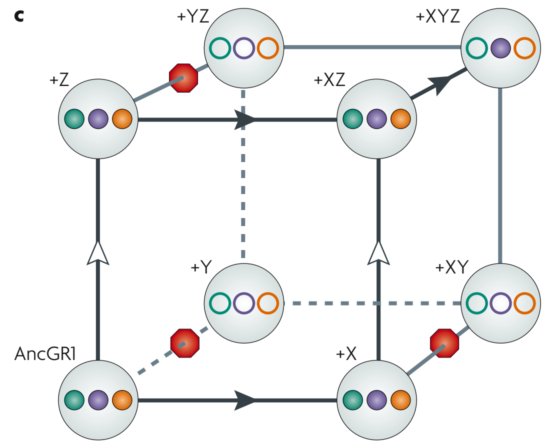


Figure 4 | **Three different views of the generation of epistasis under natural selection.** Three hypothetical adaptive landscapes are shown, in which the mean fitness of a population is a function of underlying variation at two or more loci. Topography lines display points of equal fitness, with warmer colours 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 by-product of the historical contingency built into the evolutionary process (both from the random effects of mutations and from 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[105], epistasis generates multiple adaptive peaks in the fitness landscape. If they are stable, these valleys can only be traversed by genetic drift or by tight linkage. Here, epistasis has a fundamental role in determining the direction and likelihood of any particular evolutionary change. **c** | In Gavrelet's holy landscape model[106], adaptive landscapes are rugged, as in part **b**, but the multidimensional nature of the fitness landscape leads to ridges of nearly equal fitness that populations move along, either by directional selection or by genetic drift. This is represented here by a ridge of nearly equal fitness, with end points that 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 it might affect the direction of evolutionary change, but it would not generate a barrier to change as in Wright's model.

© 2008 Macmillan Publishers Limited. All rights reserved

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 best example of this so far comes from the investigation by Ortlund *et al.*[82] into the evolution of novel function in vertebrate steroid receptors. The first step in this study was to use phylogenetic methods to reconstruct the inferred ancestral protein sequence that pre-dates the separate evolution of the mineralocorticoid and glucocorticoid steroid receptors, and to test its function[81]. It turns out that the ancestral protein is promiscuous and interacts with a variety of steroid ligands, even with ligands that were not present within the ancestral organism. Specialization therefore occurred through the evolution of a glucocorticoid-specific receptor from a more general mineralocorticoid ancestor. This process was achieved through changes at two interacting sites: serine (Ser) at position 106 replaced by proline (Pro; S106P) and leucine (Leu) at position 111 replaced by glutamine (Gln; L111Q). When these changes occur individually, S106P destroys receptor function, whereas L111Q has little functional effect. When both changes are present, however, the S106P site modifies the architecture of the protein and allows the L111Q site to form a novel hydrogen bond with cortisol — a clear case of functional epistasis (see figure part **a**; AncGR1 represents the ancestral protein, AnGR2 the modified protein). Ortlund *et al.* place proteins containing these changes into '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 further amino-acid changes (group Z) that stabilize the perturbation in protein structure that is induced by the changes in the X and Y groups. Ortlund *et al.* call the Z-group substitutions 'permissive' mutations, because they seem to have little effect on receptor function, but are a crucial step for allowing the other functional changes to occur. There is another permissive mutation, tyrosine (Tyr) at position 27 replaced by arginine (Arg; Y27R) (see figure part **b**), which precedes all of these changes, and which generates a novel cation–π interaction (that replaces a weaker hydrogen bond) that stabilizes portions of the protein that would have otherwise been destabilized by the changes described above.

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 (see figure part **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 ancestral protein (AncGR1) to the hormones aldosterone (green), cortisone (purple) and docetaxel (orange) to specificity to cortisone alone (+XYZ); filled circles represent a response to the hormone, outlined circles represent no response. 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. This figure is modified, with permission, from REF. 82 © (2007) American Association for the Advancement of Science.

© 2008 Macmillan Publishers Limited. All rights reserved

allelic difference, including those generated by induced mutations, needs to be evaluated on a case-by-case basis. Although epistasis is usually portrayed as a property of a given locus, as we have seen it is actually 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 FIG. 4) are simple metaphors of complex genetic phenomena. One of the problems in this approach is the representation of 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 that the evolutionary process is 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 localized in their effects. Kauffman has emphasized the opposite, showing that the number of valleys can rapidly increase with increasing dimensions[78]. To a large extent this is an empirical question, albeit one that is extremely difficult to address adequately.

The bottom line 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, regardless of 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 have a strong role in limiting the possible paths that evolution can take, but not in limiting its eventual outcome. Of course, this might be partly due to the fact that inaccessible evolutionary outcomes might 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 can 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 vice versa)[86] and could help to 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 complex patterns of gene regulation in localized genomic regions[87] that might be the result of similar types of evolutionary constraints. We need to look at interactions between promoters, coding genes, microRNAs, chromatin remodelling and other factors that Bateson would never have dreamed of, as being parts of epistatic networks with evolutionary dynamics that can 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 (for example, robustness, modularity and 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 (FIG. 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 elucidation of the functional nature of the interactions. 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 reductionist 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 century the concept of epistasis will be even more central to biology than it has over the past century.

© 2008 Macmillan Publishers Limited. All rights reserved

1. Phillips, P. C. The language of gene interaction. *Genetics* **149**, 1167–1171 (1998).
2. Phillips, P. C., Otto, S. P. & Whitlock, M. C. in *Epistasis and the Evolutionary Process* (eds Wolf, J. D., Brodie, E. D., III & Wade, M. J.) 20–38 (Oxford Univ. Press, Oxford, 2000).
3. Malmberg, R. L. & Mauricio, R. QTL-based evidence for the role of epistasis in evolution. *Genet. Res.* **86**, 89–95 (2005).
4. Otto, S. P. & Gerstein, A. C. Why have sex? The population genetics of sex and recombination. *Biochem. Soc. Trans.* **34**, 519–522 (2006).
5. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).
6. Carlborg, O. & Haley, C. S. Epistasis: too often neglected in complex trait studies? *Nature Rev. Genet.* **5**, 618–625 (2004).
7. Holland, J. B. Genetic architecture of complex traits in plants. *Curr. Opin. Plant Biol.* **10**, 156–161 (2007).
8. Wade, M. J. Epistasis, complex traits, and mapping genes. *Genetica* **112–113**, 59–69 (2001).
9. Azevedo, L., Suriano, G., van Asch, B., Harding, R. M. & Amorim, A. Epistatic interactions: how strong in disease and evolution? *Trends Genet.* **22**, 581–585 (2006).
10. Nadeau, J. H. Modifier genes in mice and humans. *Nature Rev. Genet.* **2**, 165–174 (2001).
11. Moore, J. H. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum. Hered.* **56**, 73–82 (2003).
12. Cordell, H. J. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **11**, 2463–2468 (2002).
A clear review of the limitations in moving from statistical estimates of epistatic effects to understanding genetic causation.
13. Demuth, J. P. & Wade, M. J. Experimental methods for measuring gene interactions. *Ann. Rev. Ecol. Evol. Systematics* **37**, 289–316 (2006).
14. Musani, S. K. *et al.* Detection of gene × gene interactions in genome-wide association studies of human population data. *Hum. Hered.* **63**, 67–84 (2007).
15. McKinney, B. A., Reif, D. M., Ritchie, M. D. & Moore, J. H. Machine learning for detecting gene–gene interactions: a review. *Appl. Bioinformatics* **5**, 77–88 (2006).
16. Marchini, J., Donnelly, P. & Cardon, L. R. Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature Genet.* **37**, 413–417 (2005).
17. Alvarez-Castro, J. M., Le Rouzic, A. & Carlborg, O. How to perform meaningful estimates of genetic effects. *PLoS Genet.* **4**, e1000062 (2008).
18. Boone, C., Bussey, H. & Andrews, B. J. Exploring genetic interactions and networks with yeast. *Nature Rev. Genet.* **8**, 437–449 (2007).
A 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.* **17**, 472–480 (2006).
20. Hansen, T. F. & Wagner, G. P. Modeling genetic architecture: a multilinear theory of gene interaction. *Theoretical Popul. Biol.* **59**, 61–86 (2001).
21. Elena, S. F. & Lenski, R. E. Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–398 (1997).
Uses randomly generated mutations in *Escherichia coli* to demonstrate that epistatic effects between loci can be highly variable and frequently cancel one another out.
22. Routman, E. J. & Cheverud, J. M. Gene effects on a quantitative trait: two-locus epistatic effects measured at microsatellite markers and at estimated QTL. *Evolution* **51**, 1654–1662 (1995).
23. Bateson, W., Saunders, E. R., Punnett, R. C. & Hurst, C. C. *Reports to the Evolution Committee of the Royal Society, Report II* (Harrison and Sons, London, 1905).
24. Beadle, G. W. Genetics and metabolism in *Neurospora*. *Physiol. Rev.* **25**, 643–663 (1945).
25. Avery, L. & Wasserman, S. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends Genet.* **8**, 312–316 (1992).
26. Huang, L. S. & Sternberg, P. W. Genetic dissection of developmental pathways. (doi:10.1895/wormbook.1.88.2) *WormBook* [online], <http://www.wormbook.org/chapters/www_epistasis.2/epistasis.html> (2005).
A comprehensive treatment of how to use classical epistasis analysis to reconstruct genetic pathways.
27. Goodwin, E. B. & Ellis, R. E. Turning clustering loops: sex determination in *Caenorhabditis elegans*. *Curr. Biol.* **12**, R111–R120 (2002).
28. Sternberg, P. W. & Horvitz, H. R. The combined action of two intercellular signaling pathways specifies three cell fates during vulval induction in *C. elegans*. *Cell* **58**, 679–693 (1989).
29. Thomas, J. H., Birnby, D. A. & Vowels, J. J. Evidence for parallel processing of sensory information controlling dauer formation in *Caenorhabditis elegans*. *Genetics* **134**, 1105–1117 (1993).
30. Tong, A. H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
A landmark paper that established the high-throughput double-deletion approach to detecting epistatic interactions.
31. Tong, A. H. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004).
32. Hartman, J. L., Garvik, B. & Hartwell, L. Principles for the buffering of genetic variation. *Science* **291**, 1001–1004 (2001).
33. Segrè, D., Deluna, A., Church, G. M. & Kishony, R. Modular epistasis in yeast metabolism. *Nature Genet.* **37**, 77–83 (2005).
34. St. Onge, R. P. *et al.* Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nature Genet.* **39**, 199–206 (2007).
References 33 and 34 show how quantitative information can be incorporated into high-throughput interaction studies to yield deeper insights into the nature of genetic networks.
35. Kroll, E. S., Hyland, K. M., Hieter, P. & Li, J. J. Establishing genetic interactions by a synthetic dosage lethality phenotype. *Genetics* **143**, 95–102 (1996).
36. Sopko, R. *et al.* Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319–330 (2006).
37. Greenspan, R. J. The flexible genome. *Nature Rev. Genet.* **2**, 383–387 (2001).
38. Lehner, B., Crombie, C., Tischler, J., Fortunato, A. & Fraser, A. G. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nature Genet.* **38**, 896–903 (2006).
39. Davierwala, A. P. *et al.* The synthetic genetic interaction spectrum of essential genes. *Nature Genet.* **37**, 1147–1152 (2005).
40. Tischler, J., Lehner, B. & Fraser, A. G. Evolutionary plasticity of genetic interaction networks. *Nature Genet.* **40**, 390–391 (2008).
41. Wong, S. L. *et al.* Combining biological networks to predict genetic interactions. *Proc. Natl Acad. Sci. USA* **101**, 15682–15687 (2004).
42. Beyer, A., Bandyopadhyay, S. & Ideker, T. Integrating physical and genetic maps: from genomes to interaction networks. *Nature Rev. Genet.* **8**, 699–710 (2007).
43. Pattin, K. A. & Moore, J. H. Exploiting the proteome to improve the genome-wide genetic analysis of epistasis in common human diseases. *Hum. Genet.* **124**, 19–29 (2008).
44. Zhu, J. *et al.* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genet.* **40**, 854–861 (2008).
Shows how interaction information from many sources can be combined to provide a more comprehensive picture of interaction networks.
45. Collins, S. R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).
46. Carlborg, O., Jacobsson, L., Ahgren, P., Siegel, P. & Andersson, L. Epistasis and the release of genetic variation during long-term selection. *Nature Genet.* **38**, 418–420 (2006).
47. Stylianou, I. M. *et al.* Quantitative trait locus analysis for obesity reveals multiple networks of interacting loci. *Mamm. Genome* **17**, 22–36 (2006).
48. Ehrenreich, I. M., Stafford, P. A. & Purugganan, M. D. The genetic architecture of shoot branching in *Arabidopsis thaliana*: a comparative assessment of candidate gene associations vs. quantitative trait locus mapping. *Genetics* **176**, 1223–1236 (2007).
49. Alvarez-Castro, J. M. & Carlborg, O. A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics* **176**, 1151–1167 (2007).
50. Cheverud, J. M. in *Epistasis and the Evolutionary Process* (eds Wolf, J., Brodie, E. D., III & Wade, M. J.) 58–81 (Oxford Univ. Press, Oxford, 2000).
51. Sambandan, D., Yamamoto, A., Fanara, J. J., Mackay, T. F. & Anholt, R. R. Dynamic genetic interactions determine odor-guided behavior in *Drosophila melanogaster*. *Genetics* **174**, 1349–1363 (2006).
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.* **115**, 429–442 (2007).
53. Rowe, H. C., Hansen, B. G., Halkier, B. A. & Kliebenstein, D. J. Biochemical networks and epistasis shape the *Arabidopsis thaliana* metabolome. *Plant Cell* **20**, 1199–1216 (2008).
54. Wolf, J. B., Leamy, L. J., Routman, E. J. & Cheverud, J. M. Epistatic pleiotropy and the genetic architecture of covariation within early and late-developing skull trait complexes in mice. *Genetics* **171**, 683–694 (2005).
55. Sinha, H., Nicholson, B. P., Steinmetz, L. M. & McCusker, J. H. Complex genetic interactions in a quantitative trait locus. *PLoS Genet.* **2**, e13 (2006).
56. Nogami, S., Ohya, Y. & Yvert, G. Genetic complexity and quantitative trait loci mapping of yeast morphological traits. *PLoS Genet.* **3**, e31 (2007).
57. Storey, J. D., Akey, J. M. & Kruglyak, L. Multiple locus linkage analysis of genomewide expression in yeast. *PLoS Biol.* **3**, e267 (2005).
58. Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436**, 701–703 (2005).
Shows how genetical genomics can be used to infer patterns of gene interaction.
59. Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).
60. Cheverud, J. M. & Routman, E. J. Epistasis and its contribution to genetic variance components. *Genetics* **139**, 1455–1461 (1995).
61. Phillips, P. C. & Johnson, N. A. The population genetics of synthetic lethals. *Genetics* **150**, 449–458 (1998).
62. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
63. Tsai, C. T. *et al.* Renin-angiotensin system gene polymorphisms and coronary artery disease in a large angiographic cohort: detection of high order gene–gene interaction. *Atherosclerosis* **195**, 172–180 (2007).
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.* **70**, 726–737 (2006).
65. Abou Jamra, R. *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.* **81**, 974–986 (2007).
66. Coutinho, A. M. *et al.* Evidence for epistasis between *SLC6A4* and *ITGB3* in autism etiology and in the determination of platelet serotonin levels. *Hum. Genet.* **121**, 243–256 (2007).
67. Gregersen, J. W. *et al.* Functional epistasis on a common MHC haplotype associated with multiple sclerosis. *Nature* **443**, 574–577 (2006).
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. *Nature Med.* **12**, 1119–1121 (2006).
69. Svejgaard, A. The immunogenetics of multiple sclerosis. *Immunogenetics* **60**, 275–286 (2008).
70. Gauderman, W. J. Sample size requirements for association studies of gene–gene interaction. *Am. J. Epidemiol.* **155**, 478–484 (2002).
71. Carlson, C. S., Eberle, M. A., Kruglyak, L. & Nickerson, D. A. Mapping complex disease loci in whole-genome association studies. *Nature* **429**, 446–452 (2004).
72. Xu, S. & Jia, Z. Genomewide analysis of epistatic effects for quantitative traits in barley. *Genetics* **175**, 1955–1963 (2007).
73. Demant, P. Cancer susceptibility in the mouse: genetics, biology and implications for human cancer. *Nature Rev. Genet.* **4**, 721–734 (2003).

© 2008 Macmillan Publishers Limited. All rights reserved

74. Jacob, F. Evolution and tinkering. *Science* **196**, 1161–1166 (1977).
75. Lynch, M. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl Acad. Sci. USA* **104** (Suppl. 1), 8597–8604 (2007).
76. Crow, J. F. How important is detecting interaction? *Behav. Brain. Sci.* **13**, 126–127 (1990).
77. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
78. Kauffman, S. A. *The Origins of Order: Self-Organisation and Selection in Evolution* (Oxford Univ. Press, New York, 1993).
79. Wu, C.-I. & Palopoli, M. F. Genetics of postmating reproductive isolation in animals. *Annu. Rev. Genet.* **27**, 283–208 (1994).
80. de Visser, J. A. *et al.* Perspective: evolution and detection of genetic robustness. *Evolution* **57**, 1959–1972 (2003).
81. Bridgham, J. T., Carroll, S. M. & Thornton, J. W. Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101 (2006).
82. Ortlund, E. A., Bridgham, J. T., Redinbo, M. R. & Thornton, J. W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544–1548 (2007).
A good example of moving between detailed functional analysis and long-term evolutionary inference.
83. Miller, S. P., Lunzer, M. & Dean, A. M. Direct demonstration of an adaptive constraint. *Science* **314**, 458–461 (2006).
84. Weinreich, D. M., Delaney, N. F., Depristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
85. Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M. & Tans, S. J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–386 (2007).
86. Karlin, S. General two locus selection models: some objectives, results and interpretations. *Theoret. Popul. Biol.* **7**, 364–398 (1975).
87. Encode Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
88. Moore, J. H. & Williams, S. M. Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. *Bioessays* **27**, 637–646 (2005).
89. Wagner, G. P., Pavlicev, M. & Cheverud, J. M. The road to modularity. *Nature Rev. Genet.* **8**, 921–931 (2007).
90. Gjuvsland, A. B., Hayes, B. J., Omholt, S. W. & Carlborg, O. Statistical epistasis is a generic feature of gene regulatory networks. *Genetics* **175**, 411–420 (2007).
91. Deutscher, D., Meilijson, I., Kupiec, M. & Ruppin, E. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nature Genet.* **38**, 993–998 (2006).
92. Jansen, R. C. Studying complex biological systems using multifactorial perturbation. *Nature Rev. Genet.* **4**, 145–151 (2003).
A perspective on how complex genetic systems can be best interrogated using multiple, rather than single, perturbations.
93. Carter, G. W. *et al.* Prediction of phenotype and gene expression for combinations of mutations. *Mol. Syst. Biol.* **3**, 96 (2007).
94. Bateson, W. *Mendel's Principles of Heredity* (Cambridge Univ. Press, Cambridge, 1909).
95. Fisher, R. A. The correlations between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433 (1918).
96. Tachida, H. & Cockerham, C. C. A building block model for quantitative genetics. *Genetics* **121**, 839–844 (1989).
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, M. W. Simultaneous stability of $D = 0$ and $D \neq 0$ for multiplicative viabilities at two loci. *Genetics* **90**, 813–825 (1978).
98. Mani, R., St. Onge, R. P., Hartman, J. L. IV, Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc. Natl Acad. Sci. USA* **105**, 3461–3466 (2008).
Shows how dependent the inference of epistasis is upon the scale of measurement.
99. Aylor, D. L. & Zeng, Z. B. From classical genetics to quantitative genetics to systems biology: modeling epistasis. *PLoS Genet.* **4**, e1000029 (2008).
100. Feldman, M. W., Otto, S. P. & Christiansen, F. B. Population genetic perspectives on the evolution of recombination. *Annu. Rev. Genet.* **30**, 261–295 (1997).
101. Bennett, D. C. & Lamoreux, M. L. The color loci of mice — a genetic century. *Pigment Cell Res.* **16**, 333–344 (2003).
102. Steiner, C. C., Weber, J. N. & Hoekstra, H. E. Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.* **5**, e219 (2007).
103. Silvers, W. *The Coat Colors of Mice: A model for mammalian gene action and interaction* (Springer, Berlin, 1979).
104. Hoekstra, H. E. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **97**, 222–234 (2006).
105. Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. 6th Int. Cong. Genet.* **1**, 356–366 (1932).
106. Gavrilets, S. *Fitness landscapes and the Origin of Species* (Princeton Univ. Press, Princeton, 2004).

Acknowledgements

This work was initiated while the author was a sabbatical visitor at the Gulbenkian Institute of Science. I gratefully acknowledge their support. I also deeply appreciate input from H. Hoekstra, S. Otto, J. Thornton and three anonymous reviewers, as well as a long-term synergistic interaction with M. Whitlock. This work was supported by a fellowship from the Guggenheim Foundation, and by grants from the National Institutes of Health and the National Science Foundation.

FURTHER INFORMATION

The Phillips laboratory homepage:
http://www.uoregon.edu/~pphil/index.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2008 Macmillan Publishers Limited. All rights reserved