Chapter 3
Organization Principles in Genetic Interaction Networks

Christopher Jacobs and Daniel Segrè

Abstract Understanding how genetic modifications, individual or in combinations, affect phenotypes is a challenge common to several areas of biology, including human genetics, metabolic engineering, and evolutionary biology. Much of the complexity of how genetic modifications produce phenotypic outcomes has to do with the lack of independence, or epistasis, between different perturbations: the phenotypic effect of one perturbation depends, in general, on the genetic background of previously accumulated modifications, i.e., on the network of interactions with other perturbations. In recent years, an increasing number of high-throughput efforts, both experimental and computational, have focused on trying to unravel these genetic interaction networks. Here we provide an overview of how systems biology approaches have contributed to, and benefited from, the study of genetic interaction networks. We focus, in particular, on results pertaining to the global multilevel properties of these networks, and the connection between their modular architecture and their functional and evolutionary significance.

1 Introduction: Epistasis and Evolutionary Systems Biology

Genetic modifications underlie several important aspects of biology. It is through genetic modifications that organisms evolve. Genetic modifications are used in genetic engineering and synthetic biology to redesign and optimize cells for

C. Jacobs
Graduate Program in Bioinformatics, Boston University, Boston, MA 02215, USA
e-mail: cljacobs@bu.edu

D. Segrè (✉)
Graduate Program in Bioinformatics, Boston University, Boston, MA 02215, USA

Department of Biology and Department of Biomedical Engineering,
Boston University, Boston, MA 02215, USA
e-mail: dsegre@bu.edu

O.S. Soyer (ed.), *Evolutionary Systems Biology*, Advances in Experimental Medicine and Biology 751, DOI 10.1007/978-1-4614-3567-9_3,

practical applications [1, 8]. It is also often through targeted or systematic genetic modifications that biologists have uncovered many aspects of how biological systems work [2, 113]. While individual genetic perturbations are important by themselves, there is something fundamental about how genetic perturbations affect a system when performed in concert. These higher order effects are particularly important when the phenotype caused by multiple perturbation is different from what one may have expected based on the individual ones [42, 81]. Such deviation from expectation is generally referred to as epistasis. Two genes, alleles, or genetic perturbations displaying epistasis are also said to have a genetic interaction.

We wish to emphasize from the start that this chapter is not meant to be a comprehensive overview of the history and importance of the concept of epistasis in biology. For this purpose, the reader could consult several recent review articles [22, 41, 88, 89, 108] and books [114], in addition to classical textbooks and literature. Rather, we will focus on a specific, relatively recent direction, namely the interplay between the concept of epistasis and the approaches and viewpoints of systems biology.

The reason epistasis is been increasingly studied in conjunction with systems biology is twofold. First, epistasis is fundamentally related to systems biology through its very definition. The behavior of the system may be drastically uncharacteristic of the behaviors of the individual components. To some extent, systems biology can be seen as the study of epistasis, i.e., of nonlinear, unexpected system-level behavior arising from combinations of components working together—the whole being more, or less, than the sum of its parts. Second, the analytical and experimental high-throughput methods of systems biology are very helpful for understanding epistasis at the cellular level. As described later, some sophisticated high-throughput technologies have been specifically designed for the purpose of systematically measuring epistasis between many genes [19, 104].

Broadly speaking, the study of genetic interactions represents a unique meeting point where biological organization principles and practical applications converge (Fig. 3.1), impacting fields as diverse as functional genomics [23, 93, 104], drug development [16, 86, 116], and immunology [78, 87].

Epistasis plays also a crucial role in evolutionary biology. An abundant literature in population genetics has been dedicated to quantitatively understanding epistasis in natural populations [114]. Epistasis affects the topology and jaggedness of the fitness landscape [90, 110] and therefore the rate and properties of adaptation. Sexual reproduction, still a perplexing phenomenon in evolutionary biology, may have evolved as a method to purge genomes of mutations through recombination [65] in response to strong deleterious epistasis between loci [4], though this idea has been the subject of debate [66, 73].

While different specific definitions and metrics for epistasis have been proposed in different contexts [22, 89], the intuitive idea of epistasis as a deviation from a null expected behavior is common to different fields, and constitutes an interesting bridge between systems biology and evolutionary biology.

In the upcoming sections, we will explore in detail some of the concepts we have just outlined. First, we will provide an overview of how epistasis may substantially differ depending on the types of perturbations performed, on the phenotype

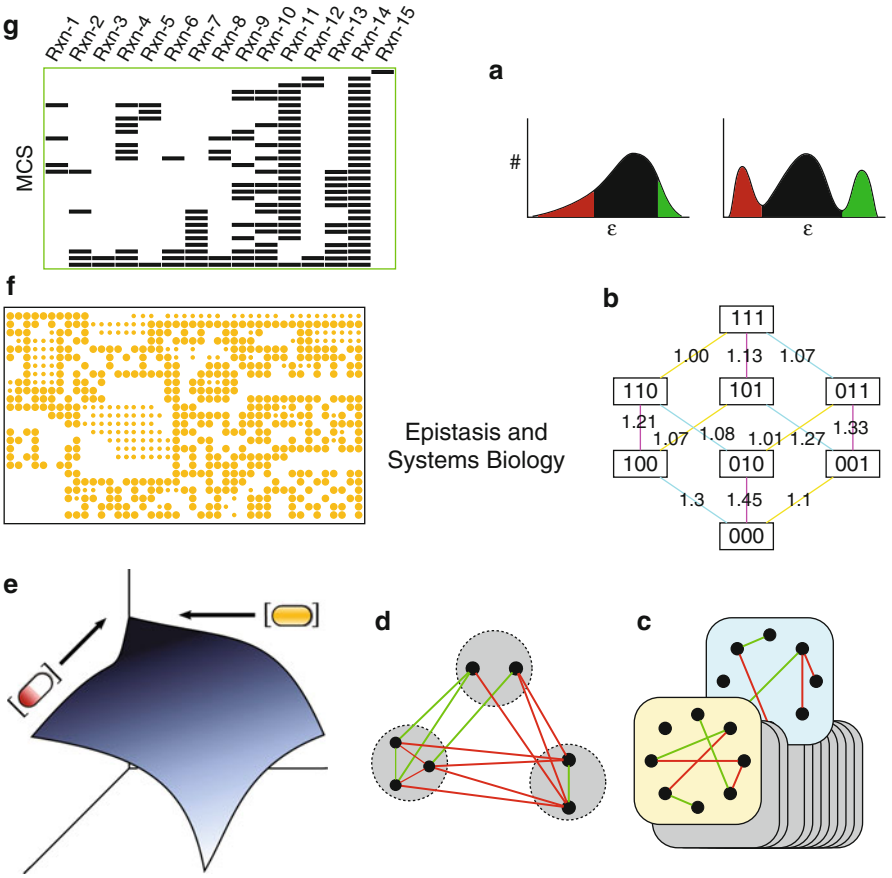

Fig. 3.1 Research areas and questions at the interface between epistasis and systems biology: **(a)** The distribution of genetic interactions between several alleles has been the subject of substantial research, largely due to its possible evolutionary implications. The definition and quantification of non-epistatic (black), synergistic (red), and antagonistic (green) effects depends in general on the null model used (e.g. multiplicative), on the type of mutations (beneficial/deleterious) and on cutoffs in the distribution. **(b)** Laboratory evolution experiments allow one to identify beneficial mutations occurring during adaptation. Epistasis (in this case antagonistic, or diminishing returns) can then be estimated by measuring fitness for all possible combinations of alleles (represented here as 3-letter strings). **(c)** Epistasis can be measured or predicted relative to any measurable trait. Hence, one can talk about multi-phenotype epistatic networks. Networks obtained relative to different phenotypes can show different patterns of antagonistic (green) and synergistic (red) interactions. **(d)** Epistatic networks can be analyzed using unsupervised clustering into monochromatically interacting modules, i.e. such that all edges between any two clusters are all of the same color. **(e)** Epistasis can be studied between drugs, in addition to genetic perturbations. Combinations of drugs in different doses give rise to drug–drug interaction landscapes. **(f)** Epistasis can be measured through high throughput assays, such as epistatic miniarrays, through which vast numbers of single- and double-deletion mutant strains can be grown in parallel, and assayed for colony size (yellow dots). **(g)** The approach of minimal cut sets (MCS) can be used to find sets (rows) of metabolic network reactions (columns) whose concurrent deletion will cause a drastic change in a specific metabolic flux phenotype, giving rise to what has been named deep epistasis

observed, and on the environmental conditions of the experiment. Next, we will illustrate a standard definition of epistasis in systems biology and the ensuing types of interactions typically encountered. We will spend then a good portion of this chapter describing how the organization of epistatic interaction networks relates to functional classification of cellular components, and how this organization varies as one monitors different phenotypes, with potential evolutionary implications. Finally, drawing from recent reports of epistasis in laboratory evolution, we will discuss how one might bridge the gap between fitness-level epistasis and epistasis at lower trait levels, perhaps heading toward a global view of the genotype–phenotype mapping and its implications to evolutionary and systems biology.

1.1 Perturbations and Phenotypes

While the central concept of epistasis in systems biology—perturbations combining in unexpected ways—is common to several studies, the embedding of this concept in specific biological systems can take many different shapes. First of all, a genetic perturbation may range from a single nucleotide polymorphism (SNP) in the coding or regulatory region of a gene, to a complete deletion of the gene, or its substitution by a different allele. Also, one can focus on either naturally occurring mutations (e.g., beneficial mutations in evolutionary experiments or natural genetic variation in a population) or artificially imposed genetic modifications (such as the systematic deletion of individual genes in an organism or engineered point mutations within a protein [82]). In systems biology, epistasis is typically assessed concurrently for multiple pairs of alleles or perturbations, or, ideally, for all possible perturbations of a certain type in a given system, e.g., the deletion of all gene pairs in a microbial species. Hence, the study of genetic interactions often entails performing high-throughput experiments or computer simulations. In turn, the type of data generated with these approaches can be effectively visualized in the form of a network, where epistatic interactions of a certain type and/or above a certain threshold can be represented as links between nodes associated with individual genes.

It is important to emphasize that the response of an organism to individual perturbations carries in itself abundant biological information, e.g., about essentiality of genes under specific conditions [85, 100]. In order to estimate epistasis, it is necessary to perform all single and all double perturbations of the alleles under study, so that the deviation between the behavior expected from two individual perturbations and the phenotype of the double perturbation can be appropriately quantified. In addition to the most elementary instance of epistasis—pair-wise interactions between perturbations—one could quantify epistasis for all possible sets of three, four, or n perturbations. Even for small genomes, though, this quickly expands to a massive undertaking. For example, to test all the possible pair-wise interactions between deletions of the approximately 6,275 genes in yeast, even assuming that a pair-wise interaction is not dependent on the order of perturbation,

one would need to carry out over 19.5 million knockout experiments. Extending such a study to include all possible triplets would need on the order of 10^{10} knockout experiments.

Another crucial parameter in the definition and quantification of epistasis is the phenotype relative to which an interaction is detected. Classical work on gene deletions, as described below, focuses on growth rate phenotype, partly because it is easily measurable, and partly because of its close relationship to evolutionary fitness in microbial systems. However, this choice is somehow arbitrary, and it is legitimate to ask whether two genes interact epistatically relative to any alternative, non-fitness phenotype. Mapping genetic influences relative to alternative phenotypes is especially important for the study of human disease, where the reduced fitness of an individual is often not readily apparent and/or is directly relatable to the expression of the alternative phenotype. For example, the aberrant phenotype of Alzheimer's disease, a neurodegenerative disease causing dementia, usually only manifests in the elderly, thus its impact on human fitness is not readily apparent until beyond the ages of reproduction. Nevertheless, Combarros et al. were able to statistically investigate 100 potential gene-pair epistatic interactions related to sporadic (i.e., non-Mendelian) Alzheimer's, eventually finding that 27 of these interactions were significantly related to Alzheimer's, including a few pairs which helped reduce the risk of onset of the disease [20]. Such studies may prove to be extremely important to human health in the future, as most traits are not under the control of a single locus [11], and epistatic interactions contributing to susceptibility and resistance seem ubiquitous throughout human disease [80].

In addition to considering multiple perturbations and multiple phenotypes, one can ask how epistasis varies for multiple environmental conditions. Though the environmental impact on human disease phenotypes has been studied for a long time [14, 20, 80], only recently has the idea of environment dependency migrated to epistatic networks in computational simulations and other investigations [3, 100, 118]. Most work in this area focuses on how epistasis depends on only one of the three key variables mentioned (perturbations, phenotype, and environment), largely because of the combinatorial explosion of possibilities, though some examples exist of studies that address the interplay between different variables, e.g., perturbations and environment [58], or perturbations and phenotypes [99]. The evolutionary implications of the environmental dependence of mutational effects and epistasis are in themselves a topic of high importance, recently addressed in RNA enzyme adaptation experiments [46].

1.2 Measuring and Predicting Epistasis

For the majority of the lifetime of the term, epistasis was quantitatively deduced by deviations from the expected relative frequencies of phenotype expression [22, 41, 82, 88]. A gene X would be epistatic to a gene Y if, the presence of the dominant allele of X (X written in italics) masked the effect of both alleles of gene Y (Y/y),

that is, the phenotypic expression of either Y or y is not observable in the presence of dominant allele X, but is observable with allele x (xx only, in diploid organisms). This was the definition of an epistatic interaction first described by Bateson and Mendel [6]. Though Bateson's definition of "epistatic" was unidirectional, it was soon after modified slightly, to lose this constraint, such that two genes could be epistatic to each other [22].

For the purpose of quantitative assessment and modeling of epistasis, it is essential to define epistasis in a more formal way, beyond the identification of phenotype masking effects. In particular, this is important for many modeling applications, including epistasis in human disease where different alleles often lead not directly to disease or immunity, but rather to increased susceptibility or resistance to the disease. This requires agreeing on a definition of what it means for a gene to have an effect on a particular trait and on assumptions about gene independence.

For quantitative traits, various mathematical/statistical models of epistasis have been developed [41, 75]. As mentioned above, we will focus here on recent definitions used in functional genomics, rather than other classical definitions found in the population genetics literature. Epistasis, in this context, can be defined as the deviation from a null model, corresponding to a multiplicative law for the combination of individual effects. In other words, epistasis is defined as:

$$\varepsilon_{ij} = W_{ij} - W_i \cdot W_j, \tag{3.1}$$

where W_{ij} is a measure of the phenotype under consideration, typically fitness, and the null expectation is then given by $W_i W_j$. All values are expressed assuming a normalized wild-type fitness $W_0 = 1$. A number of alternative metrics for measuring have been used throughout the literature, including (most notably) additive models where the null expectation matches $(W_i + W_j - 1)$, models of "minimal mathematical function" where the expectation of the double mutant is equal to the minimally "fit" of the single mutants, according to some measure (usually fitness) [41, 75], as well as many variations on the above, including heterogeneity models [22], and scaled measures of ε [95] to name only a couple of examples (more examples may be found in [41]).

An epistatic interaction may be classified as either synergistic or antagonistic. *Synergistic epistasis* (sometimes aggravating epistasis) describes an interaction which is more severe, i.e., larger in magnitude, than expected. For a combination of beneficial mutations, this would mean that ε has a positive sign, i.e., the double mutant is more fit than expected. However, combinations of deleterious mutations would have negative ε: the double mutant is less fit than expected. *Antagonistic epistasis* (sometimes buffering epistasis) describes the diminished effects of a genetic interaction, with an opposite trend relative to synergistic effects. One should be aware that the terms positive and negative epistasis can be used with different meanings in the literature. In some papers (mainly dealing with deleterious mutations), positive and negative are used to indicate respectively antagonistic and synergistic epistasis [4, 105], while others (considering mostly beneficial mutations) use positive and negative in the opposite way [35, 56]. In other works positive vs.

negative epistasis refers to the sign of ε, as defined in (3.1), where negative ε would imply antagonistic epistasis between beneficial mutations and synergistic epistasis between deleterious ones. Due to this potential ambiguity, we will avoid as much as possible the use of "positive" or "negative" epistasis throughout this chapter.

In addition to synergistic and antagonistic epistasis, it is possible to encounter cases in which not only the magnitude, but the sign (beneficial/deleterious) of a mutation changes based on the genetic background. For example, one could have deleterious effects for individual mutations ($W_i < W_0$, $W_j < W_0$), but a beneficial effect for the double mutation ($W_{ij} > W_0$). This type of epistasis, which has been named *sign epistasis* [110], may play a particularly significant role in adaptation, because it is a necessary precondition to the multi-peaked fitness landscapes [90], which force organisms to potentially go through decreased fitness (or wait for alternative phenotype-altering environmental conditions) in order to reach higher peaks.

The availability of robotics and parallelization of experimental assays made it possible to measure epistasis for a large number of genetic perturbations. Charles Boone's group began the daunting task of mapping complete epistatic interaction networks for an organism by focusing on a particular form of extreme synergistic deleterious epistasis known as synthetic sick/lethal, or SSL in baker's yeast. SSL double mutants are dead/nongrowing mutants resulting from the crossing of relatively healthy single mutants. Tong et al. introduced a new experimental methodology called the synthetic genetic array (SGA) to test SSL double mutants in a high-throughput manner in their yeast strains [103, 104]. The SGA method was later expanded upon to form E-MAPs (Fig. 3.1f), epistatic miniarray profiles [93]. E-MAPs are advantageous because they provide quantitative data on growth rate differences (based on colony size), which in turn allow both antagonistic and synergistic interactions to be observed, using a metric analogous to (3.1).

In parallel to experimental high-throughput technologies for detecting epistasis, computational biology has been used to explore the patterns and nature of epistasis using large-scale models of biological systems, often venturing into *in silico* experiments at the edge of, or beyond, experimental feasibility. In particular, the advent of whole-genome reconstructions of metabolic networks, such as the ones for *Escherichia coli* and yeast [47, 84], has made it possible to easily perform systematic and comprehensive computational screens of all possible single and double metabolic enzyme gene deletion phenotypes, producing predictions of large genetic interaction maps. One approach that has now been amply used in this context is the framework of stoichiometric constraint-based models of metabolic networks, most notably flux balance analysis (*FBA*). FBA is used to predict growth rate and metabolic fluxes (steady state rates) within networks that encompass the whole set of metabolic reactions known to be possible in a given organism (hence "genome-scale") [83]. For a more comprehensive introduction to flux balance modeling, we refer the reader to available literature (e.g., [32, 52, 83, 94]). However, we wish to stress here the fundamental assumptions behind FBA, as well as some of its limitations. FBA is based on two key simplifying assumptions. The first is that the metabolic network under study is at steady state, i.e., metabolite concentrations stay

constant over time. While this is not true for individual cells, it is often a sensible assumption for populations of cells kept under stable conditions (e.g., bacteria in a continuous flow bioreactor). The second main assumption of FBA is that the system is operating close to a set of fluxes that makes it optimal for a given task (the objective function). FBA is therefore implemented as an optimization problem that identifies the optimal flux distribution, while obeying the mass-balance constraints of steady state and the constraints imposed by the available nutrients. This problem can be efficiently solved using linear programming. For microbial systems, the maximization of biomass production has been often used as an objective function. FBA has been used to adequately predict the growth rate and byproduct secretion rates in *E. coli* [31, 106] as well as the essentiality of metabolic genes under several growth conditions [37]. Minimization of metabolic adjustment (MOMA), a variant of FBA, has been introduced to provide an alternative to the unrealistic assumption that mutant strains should be able to maximize their growth rate upon the perturbation [94]. Instead, MOMA assumes that the internal control circuitry of the cell will tend to maintain the cell close to the flux state of the wild type, compatibly with the new constraints imposed by the deletion [39, 98].

Because of their high computational efficiency—a single FBA/MOMA calculation may take less than 0.1 s—both of these methods have been widely used in large-scale perturbation studies [44, 85, 91], including predictions of epistatic interaction maps [95, 99]. Briefly, one can use FBA as the computational analogue of a high-throughput growth-rate assay, by systematically computing the effects of single and double gene deletions in a given model organism. Then, one can use (3.1), or variants thereof, to compute deviations from the multiplicative expectation. This type of analysis has been performed first in *Sacchromyces cerevisiae*, for which highly curated and tested stoichiometric reconstructions have been published in recent years [29, 47, 79].

It is important to mention that while both experimental and computational studies can evaluate growth rates and epistasis based on the multiplicative null model, a potentially thorny issue is the definition of the point beyond which a genetic interaction deviates far enough from the null model to be classified as an epistatic interaction. We will not delve into this issue in this chapter, but point the reader to relevant discussions [41, 75].

2 Modularity in Interaction Networks

As is often the case, the analysis of complex biological networks poses difficult computational and interpretational challenges. Genetic networks are no exception: they form graphs containing hundreds or thousands of nodes (genes) and interactions (epistatic links). One useful approach for understanding the biological significance of complex networks has been to organize the nodes into appropriately defined modules—self-contained units sharing common attributes—which underlie

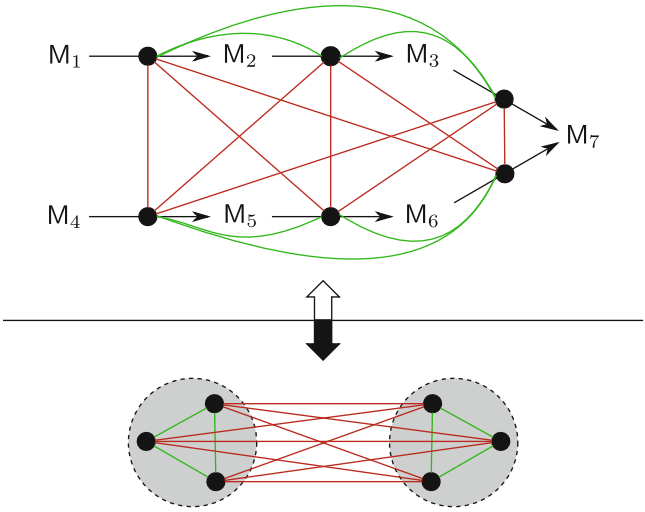

Fig. 3.2 From metabolic pathways (top) to epistatic modules (bottom). In this toy example we show how genetic interaction modules can be related to a segment of a typical metabolic network, where either of two precursor metabolites (M_1 and M_4), but not both, are required for the production of a subsequent essential metabolite (M_7). Interactions within either of the parallel pathways are antagonistic (green), because the loss of a single edge along the pathway is sufficient for the entire pathway to become defunct, thus subsequent deletions have no further impact on fitness. Interactions between pathways are synergistic (red) because even though single deletions in either pathway may be only mildly deleterious, the loss of both genes is lethal. Given that all edges between modules are of a single color, this type of hierarchical organization is named "monochromatic clustering" (see section on Hierarchy of monochromatic modules for more details)

the functional hierarchies of biology [45]. Note that a distinction has been suggested between pathways, a (usually linear) chain of information flow through a network, and modules, which do not necessarily imply a notion of information flow [15]. Despite the name, genetic interactions are not real physical interactions between genes, but rather conceptual links related to the way the system responds to their joint perturbation. Hence, according to the above definition, we expect genetic networks to form modules rather than pathways.

Functional gene modules (or simply *modules*) in epistatic networks arise from the idea that nodes (i.e. genes) have some functional relationship to one another not only if they are directly interacting, but also if their patterns of interactions with other genes display certain regularities (Fig. 3.2), e.g., if they share common neighbors. In this sense, epistatic networks can be clustered into modules using criteria and approaches similar to those implemented for clustering protein–protein interaction networks [55, 77]. Most notably, modules may be defined either as a result of enrichment of edges between the member nodes (within-module) or as a consequence of shared interactions between member nodes and nodes of distant modules (between-module).

Based on the two principles of within-group and between-group clustering, several researchers have proposed clustering schemes and applied them to different datasets in order to understand the nature of modules in epistatic networks. The SSL interaction networks generated by Tong et al. were first clustered within-group by the overlap of interactions between the first and second gene deletions [103, 104]. Segrè et al. found that FBA-generated epistasis data formed hierarchies of pathway-related modules when clustered with respect to their between-group connectivity and monochromaticity, a concept we will explore further in the next section [95]. Costanzo et al. expanded the work of the previous two studies by describing multiple types of monochromaticity in the largest yeast epistasis dataset available so far [23, 24]. Lehár et al. investigated the role of monochromaticity as an agent for selectivity within drug–drug interaction networks [69]. Guo et al. combined previous data on gene–gene interactions with gene–environment and gene–drug interaction data in their description of a recursive expectation-maximization clustering algorithm they ultimately use as a hypothesis-generating tool for investigations into the nature of robustness in cellular processes [43]. In this section, we will first describe in some detail the idea of monochromaticity in genetic networks, and then summarize some large-scale epistasis measurement efforts that corroborated the relevance and utility of this concept.

2.1 Hierarchies of Monochromatic Modules

One of the most surprising outcomes of the analysis of the genetic interaction networks predicted with flux balance modeling for yeast metabolism was the discovery of *monochromaticity* [95] (see example in Fig. 3.2). To understand the concept of monochromaticity, it is useful to recall some aspects of how epistatic interaction networks are computed through FBA. A very general property of the solutions to an FBA problem (upon maximization of the biomass production flux) is that any new constraint can only decrease the predicted growth rate. Hence, in FBA calculations, all epistatic effects occur necessarily among deleterious mutations, and synergism/antagonism refers to growth rates that are respectively smaller or larger than expected based on individual perturbations. Hence, if we draw links between epistatic gene pairs in a metabolic network and color-code them according to their class (synergistic/antagonistic), the result is a network connected by edges of two colors (conventionally red for synergistic, green for antagonistic). Upon performing a standard agglomerative hierarchical clustering algorithm, the color of the edges can be taken into account by requiring that, at every step in the clustering process, two genes (or sets of genes) can combine into a new set only if they do not interact in different colors with any other node or sets of nodes (Fig. 3.3). If this property was satisfied for a genetic interaction network, this would imply that, at any level in the hierarchy, modules would interact with each other with only one color. Indeed it was found that for the metabolic network of *S. cerevisiae* [36], the FBA-computed genetic network satisfies the property of *monochromatic*

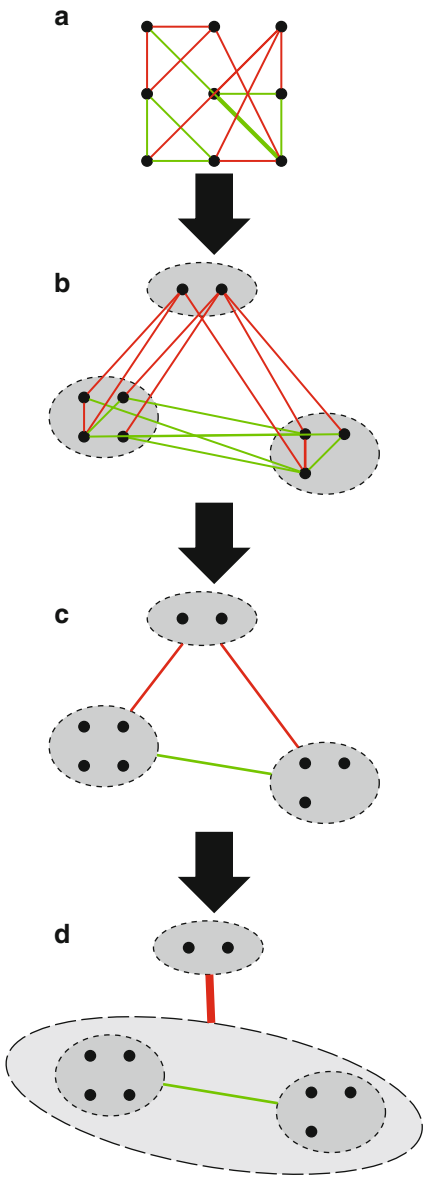

Fig. 3.3 Building hierarchies of monochromatic modules: **(a)** The epistatic interaction network is first expressed as a bi-color graph according to interaction type: red for synergistic interactions, green for antagonistic. The network is then clustered into monochromatically pure modules **(b)** using the Prism algorithm [95] (hierarchical agglomerative algorithm accounting for edge color). Groups of nodes are replaced as meta-nodes representing the entire module and **(c)** the links between module-nodes are replaced with a single representative edge between meta-nodes, before **(d)** the Prism algorithm completes the hierarchy of modules interacting monochromatically

clusterability [95]. This coherence (or monochromaticity) of interactions between modules allows one to define epistasis as a property of modules, in addition to a property of genes. Modules in metabolic networks display stronger coherent types when epistatic interactions match well against known metabolic pathways. For example, numerous genes belonging to the fermentatory pathway interact synergistically with genes belonging to respiration. The interpretation, in this case,

is that these two major energy-transducing pathways play related functional roles and cannot be simultaneously impaired without serious consequences for the cell.

It is interesting to observe that the monochromatic clusterability of the FBA-produced genetic network is not easily satisfied by random networks. In fact, the odds that a random network would be monochromatically clusterable are extremely small. In a small network, it is enough to swap a single edge color to change a monochromatically clusterable network into a non-clusterable network.

From this example of hierarchical modularity in yeast metabolism, we can see how system level properties may arise naturally from interactions at the gene level, which will be an important concept in the next sections.

2.2 Modularity and Monochromaticity in Experimental Data

While FBA-based phenotype predictions for single gene deletions can reach surprising accuracy, it is not obvious, a priori, whether properties of genetic networks discovered in silico should be expected to hold also for experimentally measured networks. In other words, is monochromatic modularity simply a theoretical construct? The idea that clustering methods would be useful to define modules of functionally related genes was already present in the early work on mapping SSL interactions in yeast [103]. The subsequent papers on SGA analysis and E-MAPs by Tong et al. [104], Schuldiner et al. [93], and Collins et al. [19] had increased focus on clustering the interaction networks resultant from their high-throughput experiments. These works mostly focused on clustering around enrichment of epistatic interactions within group. Beginning with the E-MAP data, the Boone and Weissmann groups and others have increasingly examined the role of between-group interactions, including a search for monochromaticity. Constazo et al. observed monochromatic modules of interactions across several cellular processes [23,24], e.g., metabolism and posttranscriptional modifications, and based on their observations, were able to suggest novel functional annotations for some genes (e.g., for PAR32 and SGT2) and to explain the relationship between the urmylation pathway (posttranslational modification) and elongator complex (transcription). More recently, Szappanos et al. imposed novel experimental knowledge on-top of FBA-derived epistatic interaction predictions, whereupon they found that gene dispensability can be related to degree of synergistic deleterious interactions participated in a property which itself is driven by pleiotropy [102].

The broad concept of monochromatic clustering of genetic interactions is becoming increasingly valuable as a tool for refining our understanding of cellular organization. For example, Bandyopadhyay et al. combined E-MAP data and computational predictions of epistasis with TAP-MS (tandem affinity purification followed by mass spectrometry) data, identifying proteins acting within complexes [5]. By doing so they were able to improve predictions of functionally related proteins and protein subunits, which they used to construct a functional map of

91 protein complexes involved in chromosomal architecture. This map led to the discovery of several previously uncharacterized complexes and complex subunits.

Hierarchical modularity has also been applied to classifying drug–drug interactions. Yeh et al. have applied the concept of hierarchical monochromatic clustering to epistatic networks between pairs of drugs [116, 117]. These clusters also map well into classes based on their putative functions, with the exception of drugs affecting the two subunits of ribosomes, which form two classes of protein synthesis inhibitors (PSIs). The separation of PSIs between functional classes was not something which had been noted before, and indeed many of the class–class interactions between drugs had not been well characterized. In related drug–drug interaction screens and clustering, Lehár et al. showed how some combinations of drugs may increase their selectivity [69], a reversal of what is commonly feared by prescribing multiple drugs.

These examples demonstrate how epistasis constitutes an organizing principle for the hierarchy of biological networks, with important practical applications. A fascinating, mostly unanswered question is how evolutionary adaptation gives rise to this unique architecture, and—conversely—whether and how this hierarchical modular organization imposes constraints on evolutionary trajectories.

3 Epistasis and Robustness Relative to Multiple Quantitative Traits

Epistasis, in the context of systems biology and evolutionary biology of populations, is often interpreted as the mutual dependence of genetic modifications in their impact on fitness. Interestingly, however, in other contexts—most notably in the study of human disease—researchers care about epistasis insofar as it affects alternative (i.e. non-fitness) measurable traits, such as the predisposition to a genetic disease [14, 33], or the level of metabolites in the blood, bone, etc. [96]. The effect of epistasis on non-fitness phenotypes plays also an important role in metabolic engineering, where the concurrent tinkering with multiple genes is aimed at increasing a practically important phenotype, typically the production of specific industrially or medically important molecules [54,57,64]. Might non-fitness phenotypes play an important role also in systems and evolutionary biology?

Genes, and thus epistasis, ultimately act upon fitness by acting on the intermediate phenotypes which comprise fitness. Hence, there are several reasons why alternative phenotypes are relevant to systems and evolutionary biology: (1) Even if one observes epistasis relative to fitness, it is unclear whether this is the result of epistasis relative to some specific trait (e.g. nutrient uptake rate) propagating all the way to fitness, or the outcome of interference amongst several traits; (2) Knowing that two genes are interacting relative to fitness does not provide much information on the underlying molecular mechanism for this interaction; (3) The existence of epistasis relative to various intracellular traits (e.g. size of a given

metabolite pool) would imply that simultaneous changes in multiple genes could nonlinearly alter cellular dynamics, posing new questions on the evolutionary and regulatory constraints on cellular organization.

Research on polygenic quantitative trait loci (QTLs) has been concerned with epistasis relative to non-fitness phenotypes for many years. Such alternative phenotypes may include any quantifiable trait, including metabolic abundance [30], penetrance for disease [48, 115], and several plant-related traits including the two just mentioned [61, 72, 119]. Most relevant to systems biology, largely because of the high-throughput nature, are gene expression QTLs, also referred to as eQTLs. Mapping eQTLs in clonal yeast populations has removed some of the complexity in identifying causal loci, allowing Brem et al. to trace the global expression patterns of over 1,500 yeast genes to causative loci [11, 12]. Epistasis plays a major role in this study as we will see below in the sub-section on robustness.

Taking a system-level perspective, gene expression quantitative traits are one of many possible phenotypes quantitatively measurable in the cell. However, outside of fitness and expression, large datasets suitable for assessing the degree and nature of epistasis relative to multiple quantitative phenotypes are not readily available. This is why genome-scale models of biological networks can be helpful for a preliminary assessment of such multi-phenotype maps.

3.1 Phenotype-Specific Epistasis in Metabolic Networks

In flux balance modeling, each calculation produces, in addition to growth rate, a prediction of all the metabolic fluxes in the cell. This fact offers the opportunity to utilize these fluxes as quantitative traits relative to which epistasis can be estimated. Snitkin and Segrè used flux balance modeling (specifically, MOMA) to compute the entire genetic interaction map for all double mutants in the yeast model with respect to all metabolic flux phenotypes [99]. As before, interactions could be largely classified into antagonistic and synergistic relationships between gene pairs. It is worth mentioning that, in this case, sign epistasis could occur as well, due to the fact that flux phenotypes may increase or decrease upon perturbations, whereas, in growth-optimized FBA simulations, the growth rate can only decrease upon perturbation.

A key question one can ask about these genetic interaction networks is how similar their connectivity is relative to different flux phenotypes. The model calculations predict that these networks can be quite different, reflecting the fact that different fluxes highlight different regions of the metabolic chart (see below). This can also be expressed in terms of the number of new interactions that each phenotype highlights relative to other phenotypes. Across all phenotypes, more than 2,200 unique epistatic interactions were observed, far more than can be found for fitness or any of the alternative phenotypes alone (see Fig. 4 in [99]). Approximately 80 out of 300 different phenotypes are required to capture all unique epistatic interactions.

One should keep in mind that these numbers depend on the statistical cutoff used to determine epistasis, and should not be interpreted necessarily as universal quantities.

A specific consequence of the diversity of epistatic maps relative to different phenotypes is that genes can change the sign of interaction depending on the phenotype monitored. Similar to (3.1), for a phenotype k, epistasis can be defined as follows:

$$\varepsilon_{ij}^{k} = W_{ij}^{k} - W_{i}^{k} \cdot W_{j}^{k}. \tag{3.2}$$

The phenotype-dependence of the sign of epistasis could then be expressed by saying that a pair of gene knockouts (i, j) could have synergistic epistasis relative to phenotypes $\{k_1, k_2, \ldots, k_q\}$, and antagonistic epistasis relative to phenotypes $\{k_{q+1}, k_{q+2}, \ldots, k_h\}$. This is indeed abundantly observed in the computationally generated flux balance predictions (see Fig. 3.3 in [99]). These predicted mixed interactions indicate that epistasis is not an absolute characteristic of gene-pairs, but should be contextualized by the phenotype being examined. To our knowledge specific instances of this phenomenon have not been documented experimentally yet. Since several metabolic fluxes (in particular uptake and secretion rates) are experimentally measurable, it should be possible to directly test many of these predictions in the future.

So far, we have mostly discussed the connectivity and phenotype-dependent sign of epistasis in multi-phenotype interaction networks. Next, we want to illustrate the biological insight that multiple phenotype maps can provide. One concept emerging from flux balance predictions of these maps is that different phenotypic readouts provide useful mechanistic insight about the interacting genes or processes, much more than growth rate alone would do. While in growth-based interaction maps the only way to relate genes to function is through clustering and modular organization (two genes interacting may be inferred to have related functions, but there is no information on what that function is), in multi-phenotype maps, knowing that two gene interact relative to a specific metabolic phenotype is in itself informative about the functional relationship between those genes. Two examples of predicted epistatic interactions not visible relative to growth rate, reported by Snitkin et al., illustrate this point. The first example, a synergistic relationship between serine biosynthesis and the genes comprising electron transport chain complex II, results in unexpectedly large secretions of succinate (which in this case can be considered as the observed phenotype). This occurs because the alternate predicted pathway for serine biosynthesis includes succinate as an additional byproduct. A further synergistic relationship between glutamate synthase and the electron transport chain results in surprisingly large secretions of glycerol. Hence, similar to monochromatic modules [95], and to environment-dependent perturbations [97], also multi-phenotype interaction maps can in principle help annotate genes with unknown functions, and infer relationships between processes.

The predicted existence of several epistatic interactions between different cellular processes relative to a multitude of metabolic phenotypes is yet to be directly tested experimentally. However, it was found that genes highly interacting with

other genes through antagonistic interactions relative to multiple phenotypes tend to evolve slower, providing indirect evidence for the value of these predictions, and the importance of these networks from the adaptive standpoint [99].

3.2 Multi-Phenotype k-Robustness in Metabolism

One of the consequences of epistasis measured across multiple traits is evolved robustness of cellular systems due to availability of alternative routes to many destinations. Here we use the term *robustness* to indicate the constancy of a particular (quantitative) trait in the face of large numbers of genetic perturbations. For example, one can think of the entire metabolic network of yeast as being robust under rich growth medium, because less than 20% of genes are essential for growth in YPD (yeast peptone dextrose, a common growth medium) [113]. Such robustness is common across several cellular subsystems [51, 71, 101, 107, 109]. It has been argued that this type of robustness may be largely due to the existence of modules whose genes are linked to each other by synergistic (i.e. aggravating) interactions [26, 109].

While throughout this chapter we have so far only dealt with pair-wise genetic interactions, it has been shown that it is not uncommon for a larger number of genes to be engaged in a single k-wise epistatic relationship. The manifestation of this phenomenon, also known as deep epistasis, gives rise to k-robustness, where multiple genes have to be deleted for a phenotypic change to be detectable [26]. One of the problems with investigating k-robustness, is that one needs to perform all combinations of k knockouts for large networks per phenotype examined. Although flux balance modeling is very useful in this context, performing exhaustive calculations beyond $k = 4$ becomes prohibitive, requiring other types of approaches to reveal the abundant k-robustness shown to exist above this k value [26, 49, 50]. In particular, the identification of k-robust models for larger values of k has been approached using minimal cut sets (MCSs, Fig. 3.1g). The idea of MCSs is to search efficiently for gene sets of arbitrary size k whose removal will result in phenotype loss, while the removal of any subset of such set would not. Initially applied to small biochemical networks [59], this approach has been adapted to genome-scale metabolic networks of *E. coli* [49] and human [50], relative to several different metabolic flux phenotypes. Notably, in these investigations, and in similar studies using in silico yeast models [26], the vast majority of k-robust modules discovered are of the highest cardinality investigated: for example, in the work of Imielinski and Belta [50], over 80% of (approximately 33,000 human) essential sets contain 9–10 redundant genes. This general trend of several traits having a high cardinality of epistasis matches well to experimental data in yeast [11, 12].

Deep epistasis and MCSs are another way in which modularity in genetic networks can be used to infer the function of genes where single knockouts fail [27, 53]. The removal of the an individual gene from a k-robust module provides context to the role that gene plays in the over-all network, both because of the

functional annotation of the other $k-1$ genes within the same module, and because the phenotype relative to which it was observed is potentially informative. Another practical use proposed for deep epistasis and robustness measures is the prediction of gene targets in pathogens, especially multidrug resistant bacteria [68].

4 Epistasis as an Organizing Principle

Computational predictions and analyses of epistasis using genome-scale models of metabolism, as well as high throughput experiments, such as SGA and E-MAP have provided snapshots of specific features of genetic interaction networks: hierarchical modularity, monochromaticity, phenotype-dependence, *k*-robustness, just to mention the ones discussed at length throughout this chapter. Several fundamental questions, however, are still open. One very important challenge is the pursuit of further understanding of the role of epistasis in evolution. While a lot of the high throughput work has been focused on the effects of epistasis between gene deletions, evolution typically involves many different scales of perturbations, from single base mutations, to whole chromosome duplication events. Another related challenge is piecing together these snapshots into a coherent view of the genotype–phenotype map, and on how evolution may have influenced (and be influenced by) its architecture and nonlinearities. In this section we will summarize some recent evidence of epistasis in evolutionary adaptation experiments, and describe how some of the conclusions drawn from these studies may suggest avenues for building an integrated model of epistasis in biological networks.

4.1 Epistasis in Evolutionary Adaptation

The recent availability of inexpensive sequencing technologies makes it possible to explore the outcome of adaptation in natural or laboratory evolution experiments. Several authors have now documented in detail the occurrence of epistasis in different settings, ranging from RNA viruses [13, 34], ribozymes [46, 70], to individual proteins [9, 111] and whole organisms [18, 21, 28, 56].

Two recent adaptive evolution experiments using *M. extorquens* and *E. coli* demonstrated the emergence of antagonistic (diminishing returns) epistasis between beneficial mutations arising during laboratory evolution [18, 56]. One of these two works, by Chou et al., analyzed evolution of a metabolically impaired *M. extorquens* strain, and identified four major beneficial mutations that provided improved fitness in the evolved strain. By introducing all possible combinations of beneficial mutations onto the ancestor's background and measuring fitness of the ensuing strains, Chou et al. were able to obtain a complete map of the fitness increase of each mutation on the any background of any possible combination of the other alleles [18]. This analysis highlighted an overall general trend of diminishing returns epistasis, a form of antagonistic epistasis whereby the fitness advantage

of a beneficial mutation decreases on top of successively more fit backgrounds (Fig. 3.1b), which is well in agreement with analogous studies [56, 67, 76]. An intriguing theoretical consideration that emerged from this study is that such diminishing returns epistasis at the level of fitness could be explained by expressing fitness (f) as the difference between two other traits, a benefit (b) and a cost (c) [25]. For the unperturbed system, fitness is then expressed as:

$$f_0 = b_0 - c_0 \tag{3.3}$$

The decomposition of fitness into benefit and cost in the Chou et al. system was largely motivated by the observation that changes in enzyme levels could tune fluxes affecting metabolic efficiency (benefit), and also alter the degree of morphological defects caused by excessive protein expression (cost). The model proposed to explain the diminishing returns trend assumes that any given mutation may independently alter both the benefit and the cost. If, for a mutation i the benefit and the cost are respectively modified by coefficients θ_i and λ_i (irrespective of previous mutations), then fitness upon an arbitrary number n of mutations can be expressed through the following generalized equation:

$$f_{ij\ldots n} = \theta_i \theta_j \ldots \theta_n b_0 - \lambda_i \lambda_j \ldots \lambda_n c_0. \tag{3.4}$$

Once b_0, c_0, and each θ and λ are inferred from the experimental data, (3.4) provides an excellent fit to all the fitness values for all possible combinations of mutations, and recapitulates the experimentally observed diminishing returns effect. Importantly, this antagonistic epistasis emerges at the level of fitness, despite the assumption that, relative to the benefit and cost traits, mutations combine multiplicatively, i.e. non-epistatically. This result underpins a fundamental property of epistatic networks, i.e. that epistasis at "high-order" phenotypes could result naturally from the interrelationship between two "low-order" phenotypes, in turn affected non-epistatically by multiple mutations [17].

While in the work by Chou et al. the decomposition of fitness into simpler traits takes the specific shape of a benefit-cost function, one should not necessarily expect that the relationships between different phenotypic traits will be obvious or intuitive. However, as explored next, we maintain that a hierarchical relationship between traits, and the emergence of epistasis when transitioning from one level of description to the one above, fit nicely with several other observations on genetic networks discussed in the previous sections.

4.2 Towards a Hierarchical Genotype–Phenotype Map

Three main principles of organization can be distilled out of the above discussions: (i) Monochromaticity: genetic interactions within and between modules tend to display coherent patterns of synergistic/antagonistic links; (ii) Phenotype-specificity:

the same pair of genes may interact with different types of interactions depending on the phenotype or trait relative to which epistasis is evaluated; (iii) Emergence of epistasis from coupling of traits: genes may display no epistasis relative to two simple traits, but could become interacting relative to a more complex trait that can be expressed as a function of the simpler traits [17]. In this final subsection we ask whether these three principles fit into a coherent view of how epistatic networks are organized.

In Fig. 3.4, we propose a possible connection between these three principles that we think captures some important aspects of genetic network organization. The two bottom panels of Fig. 3.4 display two very different genetic interaction networks resulting from measuring the two phenotypes X and Y, highlighting the phenotype-specificity of epistasis (principle (ii)). Fitness in this toy model is an arbitrary function F of the two traits X and Y. Principle (iii) suggests that it is possible for two genes to have no interaction relative to individual traits (e.g. two genes from sets S_X and S_Y), but become epistatic relative to fitness, due to the dependence of fitness on such traits, giving rise to the links between sets in the top panel. In general, the transition from low to high level could also cause the disappearance of specific epistatic links. Finally, genes that belong to sets highlighted by specific phenotypes in the lower levels will tend to cluster monochromatically (principle (i)), i.e. interact in a coherent fashion with genes that were responsive relative to a different phenotype.

4.3 Conclusions and Outlook

The subtle complexity of the multilevel relationships between different proposed organizing principles of genetic networks leaves a lot of questions unanswered. First, much of the evidence for these principles is based on partially tested computational predictions. Known limitations of flux balance methods may influence our perspective of epistasis between metabolic enzyme genes. For example, predictions of phenotypic traits and genetic interactions may be affected by the choice of the objective function [92, 98], by the presence of alternative optima in flux balance calculations [74, 94] or by the lack of explicit regulatory dynamics. Hence, we still do not really know how pervasive epistasis may end up being in real metabolic networks when measured relative to different phenotypes. Given that several genetic diseases involve the manifestation of aberrant phenotypes (typically other than fitness), the prevalence of epistasis relative to such phenotype could have important consequences on the study of human biology and diseases. In addition to the potential relevance of epistasis in genetic studies, a notable recent example of how epistasis can play a role in fighting diseases is the model-mediated discovery of a cancer-specific gene deletion, whose synthetic lethal interaction with a second perturbation makes it possible to selectively target cancer cells without affecting normal ones [38,40]. Second, if indeed so many internal degrees of freedom of a cell can be nonlinearly affected by multiple minor-effect perturbation of other variables,

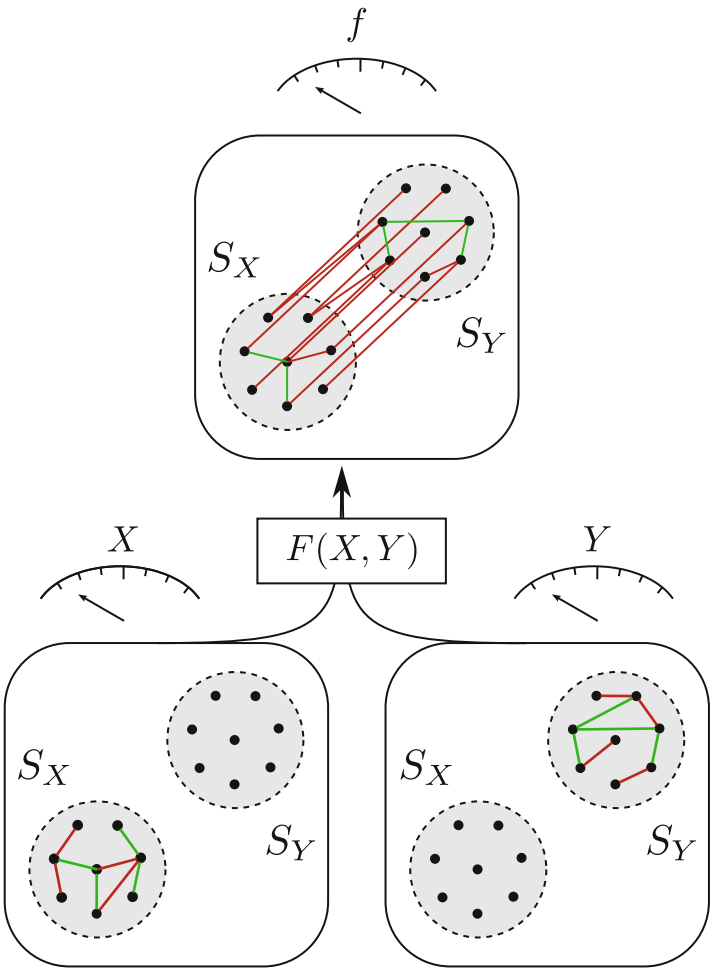

Fig. 3.4 Organizing principles of epistasis. Each panel represents the complete epistasis interaction map for a toy genome relative to the phenotypes X, Y and f. The set of genes S_X are associated with phenotype X and similarly S_Y are those genes associated with phenotype Y. The fitness phenotype, f, is dependent on the phenotypes X and Y through a function $f = F(X,Y)$. The genetic interaction map of f includes monochromatic epistasis between the sets S_X and S_Y, which could not be detected relative to either X or Y, and informs the functional relationship between X and Y

how does the cell cope? Have cells evolved, as part of their regulatory wiring, the capacity to dampen these effects, avoiding uncontrollable chaos? Or, could biological systems have embraced these epistatic effects, and learned to master them in order to control some portions of the network through subtle manipulation of more easily tunable parameters or genes? Third, it will be interesting to think whether it is possible to explain the whole hierarchy of cellular functions through multi-level traits connected by a complex, but structured web of genetic links. The

existence of *k*-robustness points to the necessity of expanding genetic interaction networks from pair-wise graphs to more complex hypergraphs [60]. Particularly important will be to try and understand how these networks have evolved, and, in turn, how they affect the rate and possibilities of evolutionary adaptation. For example, it would be interesting to explore the relationship between the robustness of metabolism relative to genetic perturbations and its robustness upon changes in environmental parameters, such as the availability of different nutrients. It is possible that the evolution of a network towards robustness to environmental uncertainty also provides robustness to single and multiple genetic perturbations under certain conditions.

Future research on epistasis will address some of the issues mentioned above through increased computational power and enhanced high-throughput experimental technologies. However, novel insights in the study of genetic interaction networks will likely stem from newly rising research directions in systems biology as well. For example, it will be interesting to explore whether nonlinearities detected at the level of population averages hold also at the single cell level, where gene expression and metabolism can be modulated by stochastic effects and cell individuality. From the mathematical perspective, several groups have started looking beyond current genome scale modeling methods, trying to incorporate thermodynamic constraints (e.g. energy balance analysis [7]), or formulate detailed mass balance models that take explicitly into account all possible macromolecules. Finally, both in the study of human biology and of microbial dynamics and evolution, we expect that a lot of new insight will come from studying the interplay of multiple cell types and microbial species. There is no reason why the synergistic and antagonistic interactions observed between genes and modules should not extend beyond the whole organism level. Stoichiometric flux balance models are already being extended from genome-scale to whole organism [10] and ecosystem level [62, 63, 112], suggesting indeed that metabolic cross-talk may play an important role in the evolution and dynamics of microbial diversity and multicellularity.

References

1. Andrianantoandro E, Basu S, Karig DK, Weiss R (2006) Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol 2:2006.0028. doi:10.1038/msb4100073
2. Arigoni F, Talabot F, Peitsch M et al (1998) A genome-based approach for the identification of essential bacterial genes. Nat Biotechnol 16:851–856. doi:10.1038/nbt0998-851
3. Arnqvist G, Dowling DK, Eady P et al (2010) Genetic architecture of metabolic rate: environment specific epistasis between mitochondrial and nuclear genes in an insect. Evolution 64:3354–3363. doi:10.1111/j.1558-5646.2010.01135.x
4. Azevedo RBR, Lohaus R, Srinivasan S et al (2006) Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature 440:87–90. doi:10.1038/nature04488
5. Bandyopadhyay S, Kelley R, Krogan NJ, Ideker T (2008) Functional maps of protein complexes from quantitative genetic interaction data. PLoS Comput Biol 4:e1000065. doi:10.1371/journal.pcbi.1000065

6. Bateson W, Mendel G (2009) Mendel's principles of heredity: a defence, with a translation of Mendel's original papers on hybridisation. Cambridge University Press, Cambridge
7. Beard DA, Babson E, Curtis E, Qian H (2004) Thermodynamic constraints for biochemical networks. J Theor Biol 228:327–333. doi:10.1016/j.jtbi.2004.01.008
8. Benner SA, Sismour AM (2005) Synthetic biology. Nat Rev Genet 6:533–543. doi:10.1038/nrg1637
9. Bershtein S, Segal M, Bekerman R et al (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. Nature 444:929–932. doi:10.1038/nature05385
10. Bordbar A, Feist AM, Usaite-Black R et al (2011) A multi-tissue type genome-scale metabolic network for analysis of whole-body physiology. BMC Syst Biol 5:180. doi:10.1186/1752-0509-5-180
11. Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci USA 102:1572–1577. doi:10.1073/pnas.0408709102
12. Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296:752–755. doi:10.1126/science.1069516
13. Burch CL, Turner PE, Hanley KA (2003) Patterns of epistasis in RNA viruses: a review of the evidence from vaccine design. J Evol Biol 16:1223–1235. doi:10.1046/j.1420-9101.2003.00632.x
14. Carlborg O, Haley CS (2004) Epistasis: too often neglected in complex trait studies? Nat Rev Genet 5:618–625. doi:10.1038/nrg1407
15. Carter GW, Rush CG, Uygun F et al (2010) A systems-biology approach to modular genetic complexity. Chaos 20:026102. doi:10.1063/1.3455183
16. Chait R, Craney A, Kishony R (2007) Antibiotic interactions that select against resistance. Nature 446:668–671. doi:10.1038/nature05685
17. Chiu H-C, Marx CJ, Segrè D Epistasis from functional dependence of fitness on underlying traits. Submitted
18. Chou H-H, Chiu H-C, Delaney NF et al (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science (New York, NY) 332:1190–1192. doi:10.1126/science.1203799
19. Collins SR, Schuldiner M, Krogan NJ, Weissman JS (2006) A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol 7:R63. doi:10.1186/gb-2006-7-7-r63
20. Combarros O, Cortina-Borja M, Smith AD, Lehmann DJ (2009) Epistasis in sporadic Alzheimer's disease. Neurobiol Aging 30:1333–1349. doi:10.1016/j.neurobiolaging.2007.11.027
21. Cooper TF, Remold SK, Lenski RE, Schneider D (2008) Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in Escherichia coli. PLoS Genet 4:e35. doi:10.1371/journal.pgen.0040035
22. Cordell HJ (2002) Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum Mol Genet 11:2463–2468. doi:10.1093/hmg/11.20.2463
23. Costanzo M, Baryshnikova A, Bellay J et al (2010) The genetic landscape of a cell. Science 327:425–431. doi:10.1126/science.1180823
24. Costanzo M, Baryshnikova A, Myers CL et al (2011) Charting the genetic interaction map of a cell. Curr Opin Biotechnol 22:66–74. doi:10.1016/j.copbio.2010.11.001
25. Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. Nature 436:588–592. doi:10.1038/nature03842
26. Deutscher D, Meilijson I, Kupiec M, Ruppin E (2006) Multiple knockout analysis of genetic robustness in the yeast metabolic network. Nat Genet 38:993–998. doi:10.1038/ng1856
27. Deutscher D, Meilijson I, Schuster S, Ruppin E (2008) Can single knockouts accurately single out gene functions? BMC Syst Biol 2:50. doi:10.1186/1752-0509-2-50
28. Draghi JA, Parsons TL, Plotkin JB (2011) Epistasis increases the rate of conditionally neutral substitution in an adapting population. Genetics 187:1139–52. doi:10.1534/genetics.110.125997

29. Duarte NC, Herrgård MJ, Palsson BØ (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 14(7):1298–1309. doi:10.1101/gr.2250904
30. Dumas M-E, Wilder SP, Bihoreau M-T et al (2007) Direct quantitative trait locus mapping of mammalian metabolic phenotypes in diabetic and normoglycemic rat models. Nat Genet 39:666–672. doi:10.1038/ng2026
31. Edwards JS, Ibarra RU, Palsson BØ (2001) In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol 19:125–130. doi:10.1038/84379
32. Edwards JS, Covert M, Palsson BØ (2002) Metabolic modelling of microbes: the flux-balance approach. Environ Microbiol 4:133–140. doi:10.1046/j.1462-2920.2002.00282.x
33. Eleftherohorinou H, Wright V, Hoggart C et al (2009) Pathway analysis of GWAS provides new insights into genetic susceptibility to 3 inflammatory diseases. PLoS ONE 4:e8068. doi:10.1371/journal.pone.0008068
34. Elena SF, Solé RV, Sardanyés J (2010) Simple genomes, complex interactions: epistasis in RNA virus. Chaos 20:026106. doi:10.1063/1.3449300
35. Eronen V-P, Lindén RO, Lindroos A et al (2010) Genome-wide scoring of positive and negative epistasis through decomposition of quantitative genetic interaction fitness matrices. PLoS ONE 5:e11611. doi:10.1371/journal.pone.0011611
36. Famili I, Fu P, Palsson BØ, Nielsen J (2003) Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. Genome Res 13(2):244–253. doi:10.1101/gr.234503.complex
37. Feist AM, Henry CS, Reed JL et al (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3:121. doi:10.1038/msb4100155
38. Folger O, Jerby L, Frezza C et al (2011) Predicting selective drug targets in cancer through metabolic networks. Mol Syst Biol 7:501. doi:10.1038/msb.2011.35
39. Förster J, Famili I, Palsson BØ, Nielsen J (2003) Large-scale evaluation of in silico gene deletions in Saccharomyces cerevisiae. OMICS 7:193–202. doi:10.1089/153623103322246584
40. Frezza C, Zheng L, Folger O et al (2011) Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. Nature 477:225–228. doi:10.1038/nature10363
41. Gao H, Granka JM, Feldman MW (2010) On the classification of epistatic interactions. Genetics 184:827–837. doi:10.1534/genetics.109.111120
42. Greenspan RJ (2001) The flexible genome. Nat Rev Genet 2:383–387. doi:10.1038/35072018
43. Guo J, Tian D, McKinney BA, Hartman JL (2010) Recursive expectation-maximization clustering: a method for identifying buffering mechanisms composed of phenomic modules. Chaos 20:026103. doi:10.1063/1.3455188
44. Harrison R, Papp B, Pál C et al (2007) Plasticity of genetic interactions in metabolic networks of yeast. Proc Natl Acad Sci USA 104:2307–2312. doi:10.1073/pnas.0607153104
45. Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. Nature 402:C47–C52. doi:10.1038/35011540
46. Hayden EJ, Ferrada E, Wagner A (2011) Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. Nature 474:92–95. doi:10.1038/nature10083
47. Herrgård MJ, Swainston N, Dobson P et al (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 26:1155–1160. doi:10.1038/nbt1492
48. Hunter KW, Crawford NPS (2008) The future of mouse QTL mapping to diagnose disease in mice in the age of whole-genome association studies. Ann Rev Genet 42:131–141. doi:10.1146/annurev.genet.42.110807.091659
49. Imielinski M, Belta C (2008) Exploiting the pathway structure of metabolism to reveal high-order epistasis. BMC Syst Biol 2:40. doi:10.1186/1752-0509-2-40
50. Imielinski M, Belta C (2010) Deep epistasis in human metabolism. Chaos 20:026104. doi:10.1063/1.3456056

51. Imielinski M, Klitgord N, Belta C (2008) Investigating the genomic basis of metabolic robustness through in silico flux analysis. IEEE Conf Decis Contr 47:793–798. doi:10.1109/cdc.2008.4739459
52. Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. Curr Opin Biotechnol 14:491–496. doi:10.1016/j.copbio.2003.08.001
53. Kaufman A, Keinan A, Meilijson I et al (2005) Quantitative analysis of genetic and neuronal multi-perturbation experiments. PLoS Comput Biol 1:e64. doi:10.1371/journal.pcbi.0010064
54. Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science (New York, NY) 330:1355–1358. doi:10.1126/science.1193990
55. Kelley R, Ideker T (2005) Systematic interpretation of genetic interactions using protein networks. Nat Biotechnol 23:561–566. doi:10.1038/nbt1096
56. Khan AI, Dinh DM, Schneider D et al (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. Science 332:1193–1196. doi:10.1126/science.1203801
57. Khosla C (2003) Metabolic engineering for drug discovery and development. Nat Rev Drug Discov 2:1019–1025. doi:10.1038/nrd1256
58. Kishony R, Leibler S (2003) Environmental stresses can alleviate the average deleterious effect of mutations. J Biol 2(2):14. doi:10.1186/1475-4924-2-14
59. Klamt S, Gilles ED (2004) Minimal cut sets in biochemical reaction networks. Bioinformatics 20:226–234. doi:10.1093/bioinformatics/btg395
60. Klamt S, Haus U-U, Theis F (2009) Hypergraphs and cellular networks. PLoS Comput Biol 5:e1000385. doi:10.1371/journal.pcbi.1000385
61. Kliebenstein D (2009) Advancing genetic theory and application by metabolic quantitative trait loci analysis. Plant Cell 21:1637–1646. doi:10.1105/tpc.109.067611
62. Klitgord N, Segrè D (2010) Environments that induce synthetic microbial ecosystems. PLoS Comput Biol 6:e1001002. doi:10.1371/journal.pcbi.1001002
63. Klitgord N, Segrè D (2011) Ecosystems biology of micrbial metabolism. Curr Opin Biotechnol 22:541–546. doi:10.1016/j.copbio.2011.04.018
64. Koffas M, Roberge C, Lee K, Stephanopoulos G (1999) Metabolic engineering. Ann Rev Biomed Eng 1:535–557. doi:10.1002/biot.201100289
65. Kondrashov AS (1988) Deleterious mutations and the evolution of sexual reproduction. Nature 336:435–440. doi:10.1038/336435a0
66. Kondrashov FA, Kondrashov AS (2001) Multidimensional epistasis and the disadvantage of sex. PNAS 98:12089–12092. doi:10.1073/pnas.211214298
67. Kvitek DJ, Sherlock G (2011) Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genet 7:e1002056. doi:10.1371/journal.pgen.1002056
68. Lehár J, Krueger AS, Zimmermann G, Borisy A (2008) High-order combination effects and biological robustness. Mol Syst Biol 4:215. doi:10.1038/msb.2008.51
69. Lehár J, Krueger AS, Avery W et al (2009) Synergistic drug combinations tend to improve therapeutically relevant selectivity. Nat Biotechnol 27:659–66. doi:10.1038/nbt.1549
70. Lehman N (2008) The molecular underpinnings of genetic phenomena. Heredity 100:6–12. doi:10.1038/sj.hdy.6801053
71. Li F, Long T, Lu Y et al (2004) The yeast cell-cycle network is robustly designed. PNAS 101:4781–4786. doi:10.1073/pnas.0305937101
72. Lisec J, Meyer RC, Steinfath M et al (2008) Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and IL populations. Plant J 53:960–972. doi:10.1111/j.1365-313X.2007.03383.x
73. MacCarthy T, Bergman A (2007) Coevolution of robustness, epistasis, and recombination favors asexual reproduction. PNAS 104:12801–12806. doi:10.1073/pnas.0705455104
74. Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng 5:264–276. doi:10.1016/j.ymben.2003.09.002
75. Mani R, St Onge RP, Hartman JL et al (2008) Defining genetic interaction. PNAS 105:3461–3466. doi:10.1073/pnas.0712255105

76. Martin G, Elena SF, Lenormand T (2007) Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nat Genet 39:555–560. doi:10.1038/ng1998
77. Michaut M, Baryshnikova A, Costanzo M et al (2011) Protein complexes are central in the yeast genetic landscape. PLoS Comput Biol 7:e1001092. doi:10.1371/journal.pcbi.1001092
78. Mitchison NA, Rose AM (2011) Epistasis: the key to understanding immunological disease? Eur J Immunol 41:2152–2154. doi:10.1002/eji.201141811
79. Mo ML, Palsson BO, Herrgård MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst Biol 3:37. doi:10.1186/1752-0509-3-37
80. Moore JH (2003) The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 56:73–82. doi:10.1159/000073735
81. Moore JH (2005) A global view of epistasis. Nat Genet 37:13–14. doi:10.1038/ng0105-13
82. Moore JH, Williams SM (2005) Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. BioEssays 27:637–646. doi:10.1002/bies.20236
83. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nat Biotechnol 28:245–248. doi:10.1038/nbt.1614
84. Orth JD, Conrad TM, Na J et al (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism – 2011. Mol Syst Biol 7:1–9. doi:10.1038/msb.2011.65
85. Papp B, Pál C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429:661–664. doi:10.1038/nature02636
86. Parsons AB, Brost RL, Ding H et al (2004) Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. Nat Biotechnol 22:62–69. doi:10.1038/nbt919
87. Penrod N, Greene CS, Granizo-MacKenzie D, Moore JH (2010) Artificial immune systems for epistasis analysis in human genetics. Lect Notes Comp Sci 6023:194–204. doi:10.1007/978-3-642-12211-8_17
88. Phillips PC (1998) The language of gene interaction. Genetics 1171:1167–1171
89. Phillips PC (2008) Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. Nature Rev Genet 9:855–867. doi:10.1038/nrg2452
90. Poelwijk FJ, Tanase-Nicola S, Kiviet DJ, Tans SJ (2011) Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. J Theor Biol 272:141–144. doi:10.1016/j.jtbi.2010.12.015
91. Raman K, Chandra N (2009) Flux balance analysis of biological systems: applications and challenges. Brief Bioinformatics 10:435–449. doi:10.1093/bib/bbp011
92. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol Syst Biol 3:119. doi:10.1038/msb4100162
93. Schuldiner M, Collins SR, Thompson NJ et al (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell 123:507–519. doi:10.1016/j.cell.2005.08.031
94. Segrè D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 99:15112–15117. doi:10.1073/pnas.232349399
95. Segrè D, Deluna A, Church GM, Kishony R (2005) Modular epistasis in yeast metabolism. Nature Genet 37:77–83. doi:10.1038/ng1489
96. Shao H, Burrage LC, Sinasac DS et al (2008) Genetic architecture of complex traits: large phenotypic effects and pervasive epistasis. PNAS 105:19910–19914. doi:10.1073/pnas.0810388105
97. Shlomi T, Herrgård MJ, Portnoy V et al (2007) Systematic condition-dependent annotation of metabolic genes. Genome Res 17:1626–1633. doi:10.1101/gr.6678707
98. Snitkin ES, Segrè D (2008) Optimality criteria for the prediction of metabolic fluxes in yeast mutants. Genome Inform 20:123–34; International Conference on Genome Informatics
99. Snitkin ES, Segrè D (2011) Epistatic Interaction Maps Relative to Multiple Metabolic Phenotypes. PLoS Genet 7(2):e1001294. doi:20.1371/journal.pgen.1001294

100. Snitkin ES, Dudley AM, Janse DM et al (2008) Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. Genome Biol 9:R140. doi:10.1186/gb-2008-9-9-r140
101. Stelling J, Sauer U, Szallasi Z et al (2004) Robustness of cellular functions. Cell 118:675–685. doi:10.1016/j.cell.2004.09.008
102. Szappanos B, Kovács K, Szamecz B et al (2011) An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat Genet 43:656–662. doi:10.1038/ng.846
103. Tong AHY, Evangelista M, Parsons AB et al (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294:2364–2368. doi:10.1126/science.1065810
104. Tong AHY, Lesage G, Bader GD et al (2004) Global mapping of the yeast genetic interaction network. Science 303:808–813. doi:10.1126/science.1091317
105. Trindade S, Sousa A, Xavier KB et al (2009) Positive epistasis drives the acquisition of multidrug resistance. PLoS Genet 5:e1000578. doi:10.1371/journal.pgen.1000578
106. Varma A, Palsson BØ (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol 60:3724–3731. doi:10.1128/AEM.00027-0254
107. de Visser JAGM, Hermisson J, Wagner GP et al (2003) Perspective: evolution and detection of genetic robustness. Evolution 57(9):1959–1972. doi:10.1111/j.0014-3820.2003.tb00377.x
108. de Visser JAGM, Cooper TF, Elena SF (2011) The causes of epistasis. Proc Roy Soc B. 278:3617–3624. doi:10.1098/rspb.2011.1537
109. Wagner A (2005) Distributed robustness versus redundancy as causes of mutational robustness. BioEssays 27:176–188. doi:10.1002/bies.20170
110. Weinreich DM, Watson RA, Chao L (2005) Sign epistasis and genetic constraint on evolutionary trajectories. Evolution 59:1165–1174. doi:10.1111/j.0014-3820.2005.tb01768.x
111. Weinreich DM, Delaney NF, DePristo MA, Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312:111. doi:10.1126/science.1123539
112. Wintermute EH, Silver PA (2010) Emergent cooperation in microbial metabolism. Mol Syst Biol 6:1–7. doi:10.1038/msb.2010.66
113. Winzeler EA, Shoemaker DD, Astromoff A et al (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science 285:901–906. doi:10.1126/science.285.5429.901
114. Wolf JB, Brodie ED, Wade JM (2000) Epistasis and the evolutionary process. Oxford University Press, New York
115. Xu C, Li Z, Xu S (2005) Joint mapping of quantitative trait loci for multiple binary characters. Genetics 169:1045–1059. doi:10.1534/genetics.103.019406
116. Yeh P, Tschumi AI, Kishony R (2006) Functional classification of drugs by properties of their pairwise interactions. Nat Genet 38:489–494. doi:10.1038/ng1755
117. Yeh PJ, Hegreness MJ, Aiden AP, Kishony R (2011) Drug interactions and the evolution of anibiotic resistance. Nat Rev Microbiol 11:330. doi:10.1038/nrmicro2133
118. You L, Yin J (2002) Dependence of epistasis on environment and mutation severity as revealed by in silico mutagenesis of phage t7. Genetics 160:1273–1281
119. Young ND (1996) QTL mapping and quantitative disease resistance in plants. Ann Rev Phytopathol 34:479–501. doi:10.1146/annurev.phyto.34.1.479