Evolved Modular Epistasis in Artificial Organisms

Sergi Valverde[1,2], Ricard V. Solé[1,2,3] and Santiago F. Elena[3,4]

[1]ICREA-Complex Systems Lab, Universitat Pompeu Fabra, Dr Aiguader 88, 08003 Barcelona, Spain
[2]Institut de Biologia Evolutiva (CSIC-UPF), Passeig Maritim de la Barceloneta, 37-49, 08003 Barcelona, Spain
[3]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501
[4]Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-UPV, 46022 València, Spain
E-mail: sergi.valverde@upf.edu

Abstract

How does complexity evolve in artificial and natural systems? A central concept within genetic systems is epistasis, namely the modulation of the effects of a given gene by one or several other genes. Epistasis is known to have an impact on many features of organisms, from recombination and sex to the ruggedness of the underlying fitness landscapes. However, the multi-scale nature of evolution and organisms makes often difficult to properly characterize epistatic interactions. Here we study the hierarchical organization of epistatic interactions between machine instructions in evolved digital organisms. We present a new quantitative approach to discover epistatic interactions that is able to capture the presence and role of groups of epistatic modules. Therefore, it thus takes into account the intrinsic nested nature of individual complexity. We found evidences of modular epistasis in avidians, with some modules having a tendency toward antagonistic epistasis while others show the opposite epistatic sign. We also found that this modular organization was positively correlated to organismal robustness.

Introduction

Genetic interactions and their impact on phenotypic traits are known to define a nonlinear mapping, which strongly affect evolutionary trajectories (Kauffman, 1993). Such nonlinear character of gene interactions is often named as *epistasis* (Van Driessche et al, 2005; Sanjuán and Elena, 2006; Collins et al., 2007, Zheng et al., 2010, Elena et al., 2010). We can understand the functional role of any component by looking at the consequences of perturbing it. Unfortunately, the above approach is limited and cannot reconstruct the functional organization of systems with ambiguous phenotype-genotype mappings. For example, we will not observe any phenotypic change if we perturb one out of two redundant components. In this context, we can extend single-perturbation experiments to double-perturbation experiments that discard redundancies explicitly.

Epistatic interactions have been used to detect functional associations between pairs of genes. Non-scaled epistasis among a pair of mutations i and j is defined as

$$\epsilon_{ij} = W_{ij} - W_i W_j \qquad (1)$$

where W_i and W_j represent the fitness values of single mutants and each entry of the matrix W_{ij} indicates the fitness value of the corresponding double mutant. Depending on the value of the above we have three different types of interactions: (1) no epistasis when $\epsilon_{i,j} = 0$, (2) synergistic epistasis when $\epsilon_{i,j} < 0$ and (3) antagonistic epistasis when $\epsilon_{i,j} > 0$.

The analysis of complex biological systems suggests that interactions between components take place between multiple scales and in the presence of feedback loops. This makes functional reconstruction a challenging and time-consuming task. In principle, the previous definition can be naturally extended to consider multiple associations, *i.e.*

$$\epsilon_{i_1,i_2,\dots,i_n} = W_{i_1,i_2,\dots,i_n} - \prod_{\mu=1}^{n} W_{i_\mu} \qquad (2)$$

but the required testing, involving multiple knockouts is, however, very costly and becomes rapidly intractable. Within the context of regulatory gene networks, it has been shown the presence of complex interactions between epistasis, network redundancy and degeneracy (Macía et al., 2012). Similarly, Lenski et al. (1999) found that epistasis was predominantly synergistic for complex digital organisms but switched to mostly antagonistic for simpler organisms.

Interestingly, the more complex organisms were also more robust against the effect of mutations than the simpler ones. The difficulties for reaching a proper understanding of the role played by epistatic interactions within complex networks calls for novel approximations. Here, we propose a new, efficient, multi-scale analysis of epistatic interactions to uncover the so-called "epistatic modules", that is, groups of related instructions and functions with similar epistatic interactions. Such approach can be useful to better understand the emergence and organization of epistasis interactions between different subcomponents of evolved organisms.

Methods

Our model organisms are digital creatures evolved within the Avida system (Ofria and Wilke, 2004). This has several
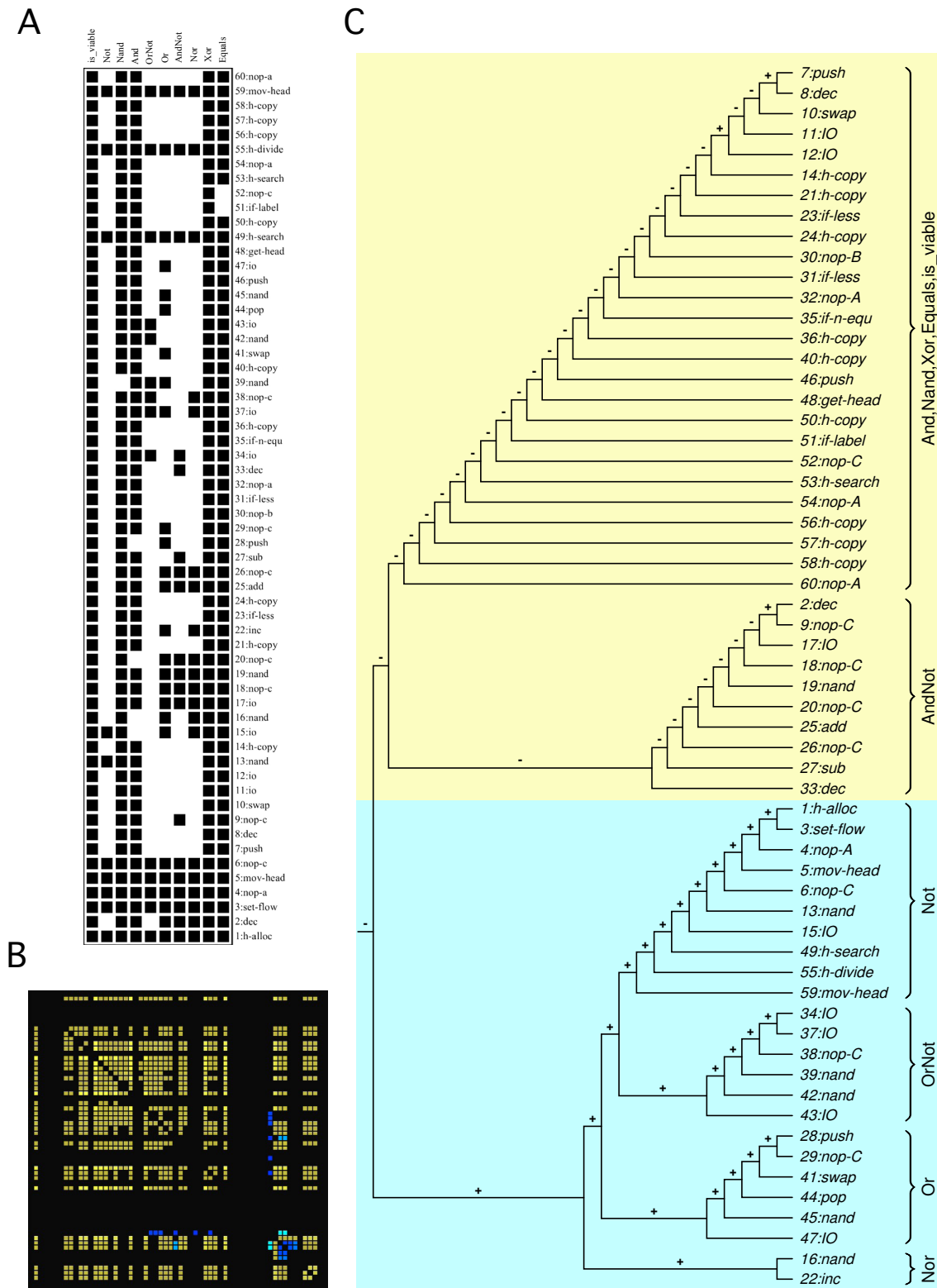
Figure 1: Analysis of functional modularity for a brittle avidian ($\langle \epsilon_\omega \rangle < 0$) with low modularity ($Q = 0.15$). (A) Task map showing the implication of each genomic instruction on the nine different tasks. (B) Heat-map illustrating the intensity of epistatic interactions between pairs of instructions in the genome. The stronger the blue, the more antagonistic (positive) epistasis; the stronger the yellow, the more synergistic (negative) epistasis. (C) Cladogram constructed from the epistasis matrix. Branches have been decorated with the average epistasis of the corresponding subtree (see text).
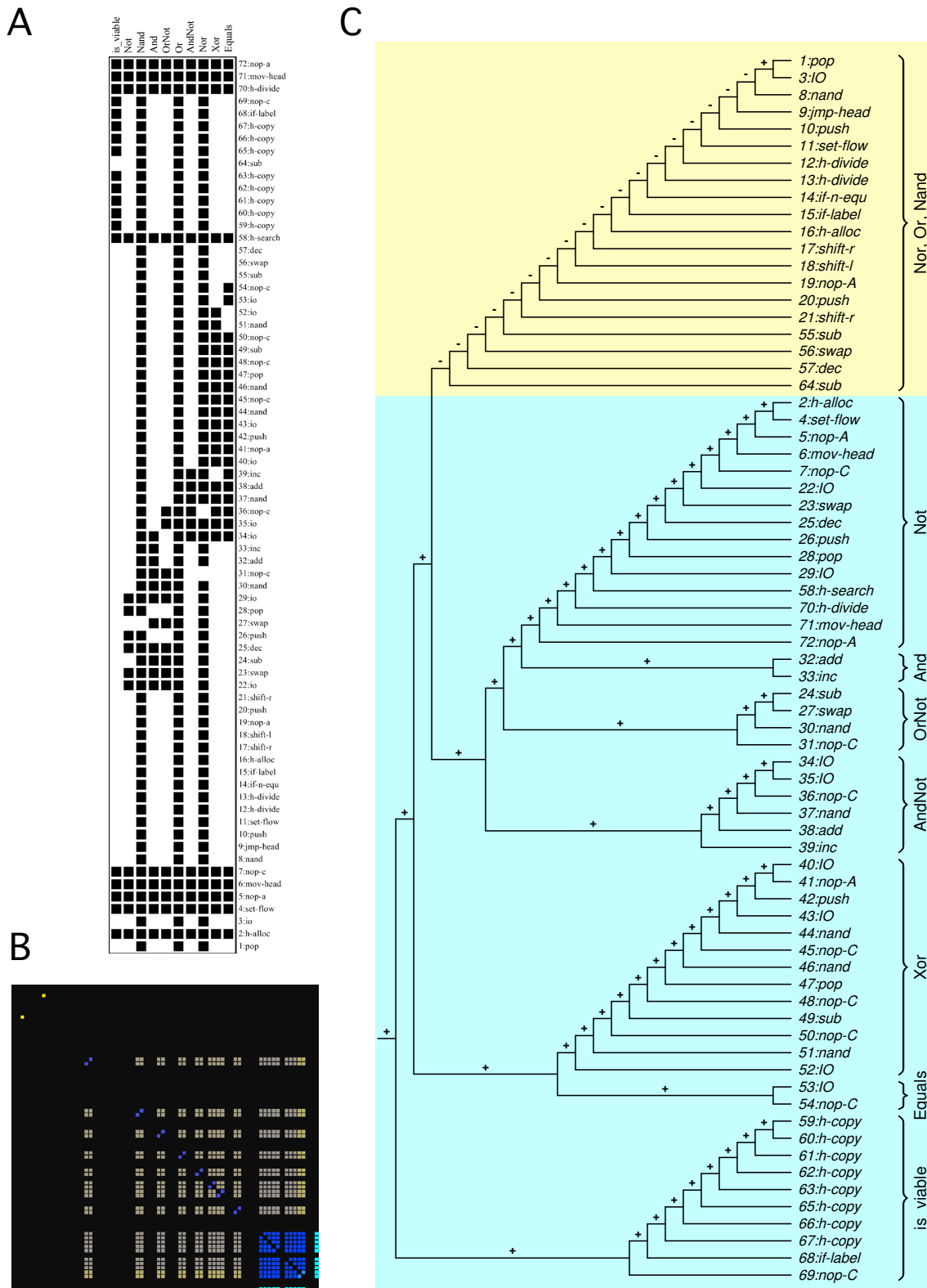
Figure 2: Analysis of functional modularity for a robust avidian ($\langle \epsilon_\omega \rangle > 0$) with high modularity ($Q = 0.21$). (A) Task map showing the implication of each genomic instruction on the nine different tasks. (B) Heat-map illustrating the intensity of epistatic interactions between pairs of instructions in the genome. The stronger the blue, the more antagonistic (positive) epistasis; the stronger the yellow, the more synergistic (negative) epistasis. (C) Cladogram constructed from the epistasis matrix. Branches have been decorated with the average epistasis of the corresponding subtree (see text).

advantages: (1) we can readily simulate many organisms in very different conditions and (2) for each artificial organism (or 'avidian') we have a clear correspondence between its genome (instructions) and the different logic tasks solved by the organism (its phenotype). An avidian consists of a CPU, a memory that stores the 'genome', registers and input/output buffers. The genome is described with a program consisting of different instructions to be interpreted by the CPU to perform different actions. The Avida system rewards any digital organism that computes a pre-defined repertoire of nine target high-level tasks. For each avidian, we also obtain a representation of the phenotype-genotype (or task) map $A_{ij} = 1$ if $j - th$ task depends on the $i - th$ instruction to be completed and $A_{ij} = 0$ if they are independent.

As it is illustrated in Fig. 1A and Fig. 2A, the same instruction can be involved in the implementation of more than one task. For example, visual inspection of the task-map in Fig. 2A indicates there are clusters of instructions with similar behavior, i.e., mutations in these instructions tend to affect the same subset of functions (e.g., instructions 22-29). Many natural and artificial networks display modular organization, that is, there are subgroups of nodes (also called modules or communities) significantly more connected between them than with the rest of nodes. Also, we can look at the pattern of connections exchanged between these modules or at the internal structure of modules (e.g., modules-within-modules).

Intuitively, we can see the modular organization as a partition of the network in distinct subparts (see below). Module detection is not an easy task because the genotype-phenotype mapping is not typically a one-to-one relationship. Here, we use the following mathematical approach to systematic module detection, in other words, disentangling the phenotype-genotype mapping. The task map A_{ij} is formally a bipartite network having two types of nodes, e.g., instructions (genotype) and functions (phenotype). This is a particular class of networks satisfying the property that there are no links between nodes of the same type, that is, interactions between functions are indirect and always mediated through, at least, one instruction. Module detection in bipartite networks is equivalent to the maximization of the so-called *modularity* (Newman and Girvan, 2004), which is an heuristic measure of the quality of any modular partition of the network:

$$Q = \frac{1}{m} \sum_{i,j} [A_{i,j} - P_{i,j}] \, \delta(g_i, g_j) \qquad (3)$$

where $m = \sum A_{i,j}$ is the total number of links, $P_{i,j} = k_i d_j / m$ is the probability that instruction i and function j are related (this takes into account the density of the task map), node i has been assigned to module g_i and $\delta(x, y) = 1$ if $x = y$ or $\delta(x, y) = 0$, otherwise. Notice that this definition of modularity is different from those previously proposed by Misevic et al. (2006) to analyze the evolution of physical and functional modularity in avidians as a result of sexual reproduction. Here, high values of Q correspond to highly modular partitions of the task map. In this case, instruction i and function j are classified in the same module so $g_i = g_j$ (and thus $\delta(g_i, g_j) = 1$) because the difference $A_{i,j} - P_{i,j} > 0$ is a large value.

The bipartite modularity algorithm finds the partition (i.e., the g_i mapping) that maximizes the Q value (Barber, 2007). From the computational point of view, the bipartite modularity algorithm is roughly equivalent to finding a hierarchical decomposition (a cladogram) of the network, that is, modular structure corresponds to a natural hierarchy of groups (and sub-groups) of instructions and functions (see Fig. 1C and Fig. 2C). We have found that the modules obtained with our algorithm have a functional meaning. Complex organisms might display a hierarchical organization of epistatic modules, where complex functions depend on simpler functions implemented at lower levels.

To better understand the relationship between modular organization and epistasis, we evaluated the sign of average epistasis for each module (and sub-module) defined in the cladogram. To do so, we compute the average epistasis for each node in the cladogram as:

$$\langle \epsilon_\omega \rangle = \frac{1}{S_\omega} \sum_{i,j \in \omega} \epsilon_{ij} \qquad (4)$$

where ω is the subset of all the instructions (e.g., the tips) in the node subtree, and ϵ_{ij} is the pairwise epistasis between tips at the lowest level of the cladogram (Eq. 1). We have implemented a new analytical tool in Avida to generate the epistasis matrices shown in Fig. 1B and Fig. 2B.

Results and Discussion

To illustrate how the algorithm works, we show the results of applying it to two avidians that were evolved to differ in their robustness against mutational perturbations (Elena and Sanjuán, 2008). In this example, the impact of changes depends on the complexity of the function associated to the mutated instructions. In aggrement with previous results (*e.g.* Lenski et al., 1999; Edlund Adami, 2004; Elena et al., 2007; Elena and Sanjuán, 2008), the more robust avidian was build in such a way that average epistasis is synergistic, whereas the brittle one shows a predominance of positive epistasis. However, our algorithm shows that the situation is not as simple as the average values may suggest, since both types of organisms, robust and brittle, are build up with modules of varying epistatic signs Indeed, brittle organisms are typically less modular although modules show epistasis of both signs (Fig. 1C). By contrast, more robust organisms are typically more modular with an abundance of antagonistic interactions, yet containing modules dominated by synergistic epistasis (Fig. 2C) .

Epistasis plays a crucial role in defining and modeling evolutionary dynamics of gene interactions and genomes. It

provides a well-defined, quantitative framework to analyze the nature and complexity of genotype-phenotype maps. Given the difficulties associated with its standard definition (Eq. 2) and under the assumption that evolved organisms involve multiple levels of nested complexity, we have proposed a network-based method to study the relationship between epistasis and modularity in artificial organisms. Such measure captures the modular nature of epistatic interactions and thus properly characterizes the internal structure of digital organisms and how they evolve and more complex, robust architectures.

Modular epistasis, that is, the situation when functional modules are constituted by genes involved epistatic interactions of a given sign, seems to be a pervasive property of biological systems (*e.g.*, Segrè et al., 2005; Costanzo et al., 2010; He et al., 2010; Xu et al. 2011). Our results suggest that selection for robustness may favor avidians which are more modular, with variance among modules in the type of epistasis they have, although showing an overall synergistic epistasis. Relaxation of the selection for robustness favors less modular organisms with an overall antagonistic epistasis, although the existing modules still may vary on the sign of epistasis. In ongoing work, we are generating extensive data resulting from the application our novel methodology to populations of avidians evolved under different genetic (robust/brittle, sexual/asexual) and environmental conditions (constant/varying environments) and will infer some generalities about the origin of genomic architecture and how it determines functional modules.

Acknowledgements

References

Barber, M. J. (2007). Modularity and community detection in bipartite networks. *Phys. Rev. E* 76:066102.

Collins, S. R., Miller, K. M., Maas, N. L., et al. (2007). Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* 446:806-810.

Costanzo, M., Baryshnikova, A., Bellay, J., et al. (2010). The genetic landscape of a cell. *Science* 327:425-431.

Edlund, J. A. and Adami, C. (2004). Evolution of robustness in digital organisms. *Artif. Life* 10:167-179.

Elena, S. F., Solé, R.V., and Sardanyés, J. (2010). Simple genomes, complex interactions: epistasis in RNA virus. *Chaos* 20: 026016.

Elena, S. F. and Sanjuán, R. (2008). The effect of genetic robustness on evolvability in digital organisms. *BMC Evol. Biol.* 8:284.

Elena, S. F., Wilke, C. O., Ofria, C., and Lenski, R. E. (2007). Effect of population size and mutation rate on the evolution of mutational robustness. *Evolution* 61:666-674.

He, X., Qian, W., Wang, Z., Li, Y. and Zhang, J. (2010). Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. *Nat. Genet.* 42:272-276.

Kauffman, S. A. (1993). *Origins of order*. Oxford University Press, New York.

Lenski, R.E., Ofria, C., Collier, T. C. and Adami, C. (1999). Genome complexity, robustness and genetic interactions in digital organisms. *Nature* 400:661-664.

Macía, J., Solé, R. V. and Elena, S. F. (2011). The causes of epistasis in genetic networks. *Evolution* 66:586-596.

Misevic. D., Ofria, C. and Lenski, R.E. (2006). Sexual reproduction reshapes the genetic architecture of digital organisms. *Proc. R. Soc. B* 273:457-464.

Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* 69:026113.

Ofria, C. and Wilke, C. O. (2004). Avida: a software platform for research in computational evolutionary biology. *Artif. Life* 10:191-229.

Sanjuán, R. and Elena, S. F. (2006). Epistasis correlates to genomic complexity. *Proc. Natl. Acad. Sci. USA* 103:14402-14405.

Segrè, D., DeLuna, A., Church, G. M. and Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nat. Genet.* 37:77-83.

Van Driessche, N., Demsar, J., Booth, E. O., Hill, P., Juvan, P. Zupan, B., Kuspa, A. and Shaulsky, G. (2005). Epistasis analysis with global transcriptional phenotypes. *Nat. Genet.* 37:471-477.

Xu, L., Jiang, H., Chen, H. and Gu, Z. (2011). Genetic architecture of growth traits revealed by global epistatic interactions. *Genome Biol. Evol.* 3:909-914.

Zheng, J., Benschop, J. J., Shales, M., Kemmeren, P., Greenblatt, J., Cagney, G., Holstege, F., Li, H. and Krogan, N. J. (2010). Epistatic relationships reveal the functional organization of yeast transcription factors. *Mol. Syst. Bio.* 6:420.