

## REPORT

# Survival of the sparsest: robust gene networks are parsimonious

Robert D Leclerc

Wagner Lab, Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA

Corresponding author. Wagner Lab, Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect Street, New Haven, CT 6520, USA.

Tel.: +1 203 687 9615; Fax: +1 203 432 3870; E-mail: robert.leclerc@yale.edu

Received 18.3.08; accepted 30.6.08

**Biological gene networks appear to be dynamically robust to mutation, stochasticity, and changes in the environment and also appear to be sparsely connected. Studies with computational models, however, have suggested that denser gene networks evolve to be more dynamically robust than sparser networks. We resolve this discrepancy by showing that misassumptions about how to measure robustness in artificial networks have inadvertently discounted the costs of network complexity. We show that when the costs of complexity are taken into account, that robustness implies a parsimonious network structure that is sparsely connected and not unnecessarily complex; and that selection will favor sparse networks when network topology is free to evolve. Because a robust system of heredity is necessary for the adaptive evolution of complex phenotypes, the maintenance of frugal network complexity is likely a crucial design constraint that underlies biological organization.**

*Molecular Systems Biology* 5 August 2008; doi:10.1038/msb.2008.52

**Subject Categories:** metabolic and regulatory networks; simulation and data analysis

**Keywords:** complexity; evolvability; gene network; robustness

This is an open-access article distributed under the terms of the Creative Commons Attribution Licence, which permits distribution and reproduction in any medium, provided the original author and source are credited. Creation of derivative works is permitted but the resulting work may be distributed only under the same or similar licence to this one. This licence does not permit commercial exploitation without specific permission.

## Introduction

The synthesis of genetics, development, and evolution remains a significant challenge for theoretical and empirical research. In response to this challenge, researchers (Wagner, 1996; Salazar-Ciudad *et al.*, 2000; Siegal and Bergman, 2002; Bergman and Siegal, 2003; Solé *et al.*, 2003; Masel, 2004; Azevedo *et al.*, 2006; Huerta-Sanchez and Durrett, 2006; Oikonomou and Cluzel, 2006; Siegal *et al.*, 2006; Ciliberti *et al.*, 2007a, b; MacCarthy and Bergman, 2007a, b; Martin and Wagner, 2008) have begun to fuse systems biology, network theory, and evolutionary theory in a new research program of evolutionary systems biology. Applying a genetic algorithm (Holland, 1992) to evolve populations of individuals that are dynamically modeled as transcriptional regulatory networks, researchers have investigated issues such as the evolution of robustness and evolvability (Wagner, 1996; Salazar-Ciudad *et al.*, 2000; Siegal and Bergman, 2002; Bergman and Siegal, 2003; Solé *et al.*, 2003; Azevedo *et al.*, 2006; Huerta-Sanchez and Durrett, 2006; Ciliberti *et al.*, 2007a, b; MacCarthy and Bergman, 2007a; Martin and Wagner, 2008), the mechanisms

of genetic assimilation (Masel, 2004), the maintenance of sex (Azevedo *et al.*, 2006; MacCarthy and Bergman, 2007a), the role of network topology (Oikonomou and Cluzel, 2006; Siegal *et al.*, 2006), and gene duplication and subfunctionalization (MacCarthy and Bergman, 2007b).

Early computational studies evolving artificial gene regulatory networks have previously determined (Wagner, 1996; Siegal and Bergman, 2002) that more densely connected networks evolve to be more robust to perturbation than sparser networks under stabilizing selection. Under stabilizing selection, robustness is selected for implicitly by the fitness function because stabilizing selection will select against deviations from the phenotypic optimal. Consequently, this will favor genes and genotypes that tend to express the optimal phenotype given factors such as environmental noise and mutation. In most natural populations, stabilizing selection operates most of the time. Because natural selection should favor more robust networks, studies are now customarily reported using more densely connected networks (Wagner, 1996; Siegal and Bergman, 2002; Bergman and Siegal, 2003; Masel, 2004; Azevedo *et al.*, 2006; Ciliberti *et al.*, 2007a, b;

**Table I** Biological networks are sparsely connected

Organism	Interactions	Genes	<i>c</i>	<i>K</i>	Secondary source	Primary source
<i>Drosophila melanogaster</i>	29	14	0.148	2.07	<i>i</i>	<i>ii</i>
<i>D. melanogaster</i>	45	25	0.072	1.8	<i>iii</i>	<i>ii</i>
Sea urchin	82	44	0.0065	1.86	<i>iii</i>	<i>iv</i>
<i>Saccharomyces cerevisiae</i>	1052	678	0.0023	1.55	<i>iii</i>	<i>v</i>
<i>S. cerevisiae</i>	3969	2341	0.0007	1.7	<i>iii</i>	<i>vi</i>
<i>S. cerevisiae</i>	106	56	0.0338	1.9	<i>vii</i>	<i>viii</i>
<i>Escherichia coli</i> <sup>a</sup>	578	423	0.0032	1.37	<i>iii</i>	<i>ix</i>
<i>Arabidopsis thaliana</i> <sup>b</sup>	18 625	6760	0.0004	2.75	—	<i>x</i>

<sup>a</sup>*N* denotes operons, not genes.

<sup>b</sup>This result was derived (statistically) by partial correlation analysis on microarrays and we suspect that this method is not precise. A pilot study with just 2000 genes found a network with  $N=820$  and  $n=828$ , which gives  $c=0.00123$  and  $K\sim 1.0$ . However, this is likely much too sparse. Moreover, although this sample should be representative for larger  $N$ , and thus  $K$  should be the same for large  $N$ , their partial correlation analysis shows a  $K=2.75$ . This method seems to give a large number of false negatives for smaller  $N$ , and according to our analysis, a large number of false positives for the larger networks. Nevertheless, a value for  $K$  where  $1\leq K\leq 2.75$  will still result in a network that is quite sparse.

The number of network interactions for a subset of an organism's genes, which were reported in various studies and databases, is shown. Interactions: the number of interactions  $n$  reported for the  $N$  genes; genes: the number of genes  $N$  that had reported interactions; column  $c$ : the connectivity density ( $c=n/N^2$ ); column  $K$ : the average number of transcriptional regulators per gene ( $K=cN$ ); secondary source: the source that reported the values; primary source: where the secondary source derived the values from. References for secondary and primary sources are shown at the bottom of the table.

Sources: (i) Serov VN, Spirov AV, Samsonova MG (1998). Graphical interface to the genetic network database GeNet. *Bioinformatics* **14**: 546–547; (ii) GeNet (<http://www.bionet.nsc.ru/bgrs/thesis/17/index.html>); (iii) Rosenfeld N and Alon U (2003). Response delays and the structure of transcription networks. *J Mol Biol* **329**: 645–654; (iv) Davidson EH *et al* (2002). A genomic regulatory network for development. *Science* **295**: 1669–1678; (v) Costanzo MC *et al* (2001). YPD<sup>TM</sup>, PombePD<sup>TM</sup> and WormPD<sup>TM</sup>: model organism volumes of the BioKnowledge<sup>TM</sup> Library, an integrated resource for protein information. *Nucleic Acids Res* **29**: 75–79; (vi) Lee TI *et al* (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**: 799–804; (vii) Kauffman S, Peterson C, Samuelsson B, Troein C (2003). Random Boolean network models and the yeast transcriptional network. *Proc Natl Acad Sci USA* **100**: 14796–14799; (viii) [http://web.wi.mit.edu/young/regulatory\\_network](http://web.wi.mit.edu/young/regulatory_network); (ix) Shen-Orr SS, Milo R, Mangan S, Alon U (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* **31**: 64–68; (x) Ma S, Gong Q, Bohnert HJ (2007). An *Arabidopsis* gene network based on the graphical Gaussian model. *Genome Res* **17**: 1614–1625.

MacCarthy and Bergman, 2007a, b; Martin and Wagner, 2008). However, real gene networks including *Escherichia coli*, yeast, *Arabidopsis*, *Drosophila*, and sea urchin appear to be robust, but are all sparsely connected (see Table I); furthermore, despite the differences in phylogeny, phenotypic complexity, life history, and number of genes all surprisingly show a mean of  $K=1.5$ – $2$  transcriptional regulators per gene.

This discrepancy between our theoretical understanding and empirical data needs to be resolved because we may be overlooking something important. Moreover, because network connectivity appears to play an important role in the network dynamics of these models, and because theoretical results are typically reported with dense networks ( $K>2$ ), these results may be inapplicable, uninformative, or misleading. Indeed, results from these studies often show that the positive results trend less favorably with decreased connectivity (Wagner, 1996; Siegal and Bergman, 2002; Azevedo *et al*, 2006; MacCarthy and Bergman, 2007a; Martin and Wagner, 2008), and where data were reported for sparse networks ( $K<2.5$ ) the theoretical conclusions were shown not to hold (Siegal and Bergman, 2002; Azevedo *et al*, 2006). This suggests that dense network connectivity may be a critical assumption for many of these studies.

Previous studies investigating the evolution of robustness (Wagner, 1996; Siegal and Bergman, 2002; Bergman and Siegal, 2003; Azevedo *et al*, 2006) measured genetic robustness by the expected effects from a single perturbation applied to a random network interaction. Because perturbations can cause the phenotype to deviate, this method effectively estimates the expected cost of a perturbation per interaction (CPPI). However, by measuring genetic robustness in this way, the addition of a frivolous interaction would seemingly increase robustness whenever its average cost of perturbation (CP) was less than the CPPI. Using this procedure, spurious

complexity is awarded credit for increasing genetic robustness simply because it reduces CPPI. Rather, if anything, spurious complexity should tend to decrease robustness because it introduces an additional target for perturbations and potentially a new channel for perturbations to propagate when acting on other interactions.

To measure network robustness more appropriately, it is necessary to critically evaluate how complexity affects the network (see Supplementary information for details and derivation). Briefly, in order for an interaction to increase robustness when added to a network, the interaction must actually decrease the CP incurred by other interactions. Simply diluting the CPPI by spreading a larger total cost over a greater number of interactions does not make a network more robust. Furthermore, when an interaction is added to a network, the costs of the additional interaction must be accounted for as the network is now exposed to a new degree of freedom for perturbations can act. Thus, the benefits of an interaction must exceed its costs to be favored by selection. Failing to account for the costs of complexity, sparser networks will be penalized for efficiency.

This suggests that genetic robustness must be measured in a way that takes into account the costs of complexity. One way to account for the costs of complexity is to sum the mean CP for each interaction, to arrive at the gross CP (GCP) for a network. In this way, although the addition of spurious interactions may decrease the CPPI, robustness would still decrease as measured by an increase in GCP.

## Results and discussion

To see how the CPPI and the GCP (see Materials and methods) are affected by connectivity density, we first specify rates by

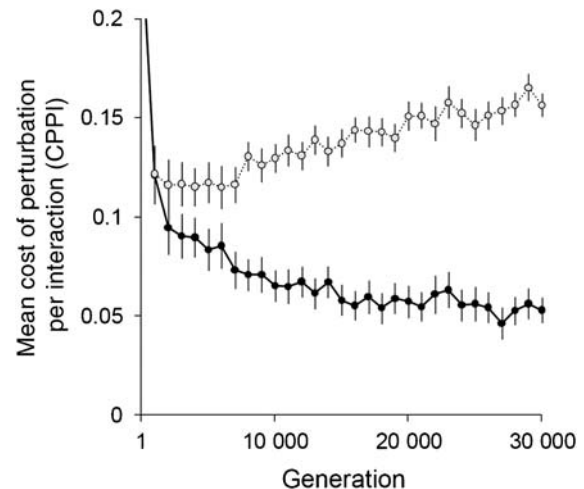
which network interactions are created ( $\alpha$ ) or destroyed ( $\phi$ ), and/or modified ( $\delta$ ) by mutation, and then apply this to a transcriptional gene regulatory model (see Materials and methods; Supplementary information) based on previous studies (Wagner, 1996; Siegal and Bergman, 2002; Bergman and Siegal, 2003; Masel, 2004; Azevedo *et al.*, 2006; Huerta-Sanchez and Durrett, 2006; Siegal *et al.*, 2006; Ciliberti *et al.*, 2007a, b; MacCarthy and Bergman, 2007a, b; Martin and Wagner, 2008). For a network of  $N$  nodes and  $n$  directed interactions (see Supplementary Figure 1), the destruction and creation of network interactions influence a network's connectivity density ( $c=n/N^2$ ; where  $K=cN$ ). In the absence of selection, connectivity density ( $c$ ) changes by:  $c'=\mu(\alpha(1-c)-\phi c)/N^2$  (see Supplementary information); where  $\mu$  is the network's mutation rate. Solving  $c(t+1)=c(t)$ , gives the expected equilibrium density:

$$\hat{c} = \frac{\alpha}{\alpha + \phi} \quad (1)$$

By setting the rates at which network interactions are added ( $\alpha$ ) or destroyed ( $\phi$ ) by mutation, we can observe how the CPPI and GCP change with connectivity density when networks are initialized far from their equilibrium density.

Both sparse ( $\hat{c}=0.1$ ;  $\alpha=0.008$ ,  $\phi=0.07$ ) and dense ( $\hat{c}=0.9$ ;  $\alpha=0.07$ ,  $\phi=0.008$ ) treatment networks of 100 replicate populations were initialized far from equilibrium ( $c_0=0.5$ ) and were evolved under stabilizing selection for 30 000 generations (see Materials and methods). Expressing the same distribution of functions at  $c(0)=0.5$ , dense treatment networks will be driven to  $c(t) \rightarrow 0.9$ , whereas the sparse treatment will be driven to  $c(t) \rightarrow 0.1$  (see Supplementary Figure 2). According to our analysis, we expect the CPPI to steadily increase as  $c(t) \rightarrow 0.1$ , and decrease as  $c(t) \rightarrow 0.9$ , but that sparser networks will be more robust at any time ( $t$ ) as measured by GCP. Fitness was scored according to an individual's phenotypic similarity to the population's founder individual; the population's ancestor whose genotype was randomly generated for each treatment and which seeded the rest of the founder population. All results were reported by sampling a single individual expressing the optimal phenotype from each replicate population. Using parameters commonly used in previous studies, for all experiments  $N=10$ ,  $\mu=0.1$ , and  $\sigma=1$  is the strength of selection, unless stated otherwise.

As expected, CPPI increases (Figure 1), whereas the GCP systematically decreases (Figure 2) with decreasing connectivity density ( $c(t) \rightarrow 0.1$ ). Conversely, as networks evolve denser connectivity ( $c(t) \rightarrow 0.9$ ) CPPI systematically decreases (Figure 1), but nevertheless, the GCP is always higher (Figure 2) compared to the sparse treatment at the same time ( $t$ ). With a smaller GCP, the sparse treatment networks simply make more efficient use of network interactions to fulfill the same function. This indicates that sparse networks are actually more robust if the costs of complexity are accounted for. If true, then evolution should seek to optimize the costs and benefits of complexity with a parsimonious network structure, a network topology that is sparsely connected and not unnecessarily complex, by seeking an optimal topological ensemble of interactions that best meets the network's functional requirements under its normal range of operating conditions.

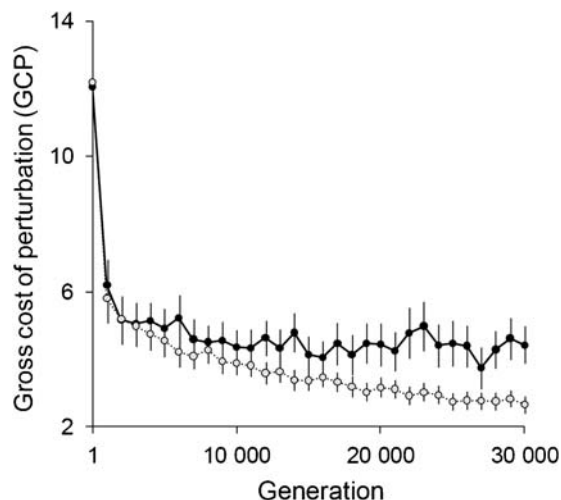


**Figure 1** Denser networks dilute the costs of perturbation over more interactions. Vertical axis shows the average cost of a perturbation per interaction (CPPI), where cost measures how much the phenotype deviates from the optimal. Selecting for optimal gene expression patterns, networks were initialized at  $c_0=0.5$  and far-from-equilibrium ( $\hat{c}=c_0 \pm 0.4$ ) for sparse ( $\hat{c}=0.1$ ;  $\alpha=0.008$ ,  $\phi=0.07$ ) and dense ( $\hat{c}=0.9$ ;  $\alpha=0.07$ ,  $\phi=0.008$ ) treatments. Plot shows evolutionary response in CPPI as networks are driven to sparser (open circles) and denser (closed circles) connectivity densities (also see Supplementary Figure 3). Data are reported as expectations and 95% confidence intervals for the mean of a single optimal individual sampled from each replicate population.

Because selection was not permitted to evolve the network 'topology' (connectivity) in previous studies—only the interaction strengths—it was never demonstrated that selection would actually favor denser networks. To test whether sparse networks will be favored by selection, we initialized replicate populations at equilibrium connectivity density by specifying appropriate values for  $\alpha$  and  $\phi$ , and then evolved populations under stabilizing selection. If sparser networks evolve greater robustness under stabilizing selection, then we expect networks to evolve connectivity density below their equilibrium value, as determined by:  $c(t) < \hat{c}$ .

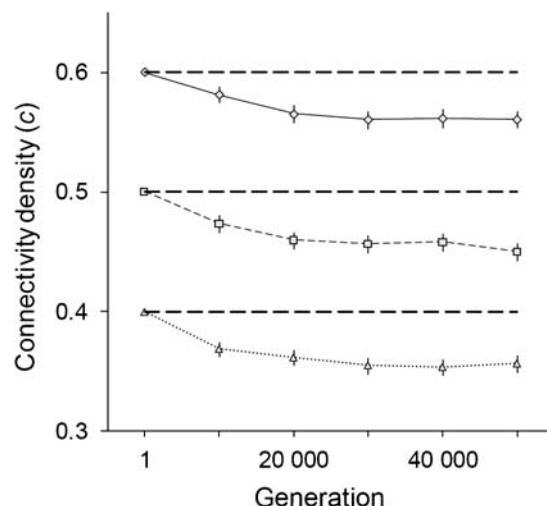
Three treatments of 100 replicate populations were classified as high ( $\hat{c}=0.6$ ;  $\alpha=0.105$ ,  $\phi=0.07$ ), intermediate ( $\hat{c}=0.5$ ;  $\alpha=0.07$ ,  $\phi=0.07$ ), or low ( $\hat{c}=0.4$ ;  $\alpha=0.07$ ,  $\phi=0.105$ ) equilibrium density, and networks were initialized to their equilibrium values ( $c_0=\hat{c}$ ). As predicted, the evolutionary response (Figure 3) shows that selection favors networks below their equilibrium density ( $c(t) < \hat{c}$ ). Similar results were also observed for asexual reproduction, and larger networks when sparse and dense networks evolved under competition (see Supplementary information), and we observed the same qualitative result for multicellular implementations and with the inclusion of developmental noise (not shown). Our analysis (see Supplementary information) indicates that this is a general result. In contrast to early studies (Wagner, 1996; Siegal and Bergman, 2002; Bergman and Siegal, 2003; Azevedo *et al.*, 2006), our results and analyses show that selection favors sparse networks, and that misassumptions about how to measure robustness in previous studies had inadvertently discounted the costs of complexity.

We have determined that selection favors the formation of sparse and minimally complex networks and that robustness



**Figure 2** Sparser networks evolve to be less costly. Vertical axis measures the gross cost of perturbation (GCP) on a network as the mean cost of perturbation per interaction (CPPI) multiplied by the number of interactions in the network. Plot shows the evolutionary response in the GCP for the treatments from Figure 2 driven toward more sparse (open circles) or dense (closed circles) connectivity densities (also see Supplementary Figure 3). For a network with  $N$  genes, the maximum cost is  $N^2$ . Sparser networks maintain a lower GCP than dense networks, even though the CPPI increases as networks become sparser. Data are reported as expectations and 95% confidence intervals for the mean of a single optimal individual sampled from each replicate population.

implies a parsimonious network. These results bring theory in line with empirical data, which show that biological gene networks are sparsely connected (Table I), having about  $K \sim 2$  transcriptional regulators per gene. These considerations challenge the theoretical conclusions of previous studies, which were reported using dense networks ( $K > 2$ ), particularly when positive results often trend less favorably with decreasing  $K$ , and have been shown to break down for  $K \leq 2.5$  (see Supplementary information). Using the fixed topology model, Azevedo *et al* (2006) reported that sex ensures its own maintenance by selecting for a negative epistatic effect (deleterious mutations will act multiplicatively). When compounded deleterious mutations are concentrated in one genotype, they will be purged according to the deterministic mutation hypothesis (Kondrashov, 1988). Despite these observations, selection for negative epistasis was not observed for sparse networks ( $K \leq 2.5$ ), whereas biological gene networks have an average of  $K = 1.5$ – $2$  transcriptional regulators per gene (Table I). Other studies have concluded that robustness could evolve under viability selection, rather than stabilizing selection (Siegal and Bergman, 2002). These results, however, were based on the analysis of dense networks ( $K \geq 4.0$ ), whereas less connected networks ( $K = 0.1444$ ) evolved minimal robustness (see Supplementary information, Supplementary Figure 6). A subsequent report (Bergman and Siegal, 2003) also suggested that arbitrary genes may buffer genetic variation and act as evolutionary capacitors, thereby providing a general mechanism to cryptically store hidden genetic variation. However, these results were reported with  $c \geq 0.75$  and  $N = 10$ ; values that are indicative of a densely connected network; nevertheless, our analyses (see



**Figure 3** Selection systematically favors networks below their equilibrium density. Selecting for optimal gene expression patterns, 100 replicate populations were initialized with network connectivity density already at equilibrium ( $c_0 = \bar{c}$ ) for high ( $c_0 = 0.6$ ;  $\alpha = 0.105$ ,  $\phi = 0.07$ ), intermediate ( $c_0 = 0.5$ ;  $\alpha = 0.07$ ,  $\phi = 0.07$ ), and/or low ( $c_0 = 0.6$ ;  $\alpha = 0.07$ ,  $\phi = 0.105$ ) treatments. Plot shows the average evolved response in connectivity density under stabilizing selection as compared to the expected equilibrium density in the absence of selection (thick dashed line). High (open diamond), intermediate (open square), and low (open triangle) systematically favor sparser-than-equilibrium networks (also see Supplementary Figure 4). Data are reported as expectations and 95% confidence intervals for the mean of a single optimal individual sampled from each replicate population.

Supplementary information) suggest that genes which buffer genetic variation should be highly connected, although the overall network should still be sparsely connected, and that a hierarchical or scale-free distribution may help optimize the costs of interactions.

Our analyses also result in several theoretical and practical implications, which may provide an alternative perspective for biological systems. The competing constraints on robustness and complexity suggest that an optimal network, for a given function, might be preordained to only a few (sparse) canonical topologies (for example, see Supplementary information); and would suggest a new interpretation for homology and developmental constraints. That is, if biological networks are optimal, then robust functional networks may reside on high adaptive peaks, divided from each other by large topological distances that are impassible by neutral evolution—except perhaps by gene duplication.

Furthermore, the observation that few optimal topologies are able to satisfy a given function would support the theory (Noman and Hitoshi, 2005) that the topologies of real biological networks can be reverse-engineered through evolutionary simulations. Previous studies support this possibility. Attempts by von Dassow *et al* (2000) to model the *Drosophila* segment polarity network from empirical data failed until two additional, previously unreported, interactions were included in their analysis. Following the addition of these factors, the network was determined to be highly robust, with the many parameters demonstrating high degrees of variation. In some instances, up to a ‘1000-fold’ variation was observed. Indeed, when a random set of values were assigned to the 48 parameters that described the network interactions, the



authors reported that  $\sim 1$  in 200 of the randomly chosen parameter sets produced the desired pattern; this lead to the conclusion that the network's topology, rather than the kinetic and biochemical details, was primarily responsible for the properties of robustness. Because the *Drosophila* segmentation network is both robust and sparse it may be possible, given sufficient computational resources, to reverse engineer this network structure with evolutionary simulations.

As selection should favor parsimonious networks, this may impose a fundamental design constraint that also drives the evolution of epigenetic processes, systems, and strategies to effectively maintain frugal network complexity despite increases in genomic complexity. Consistent with this hypothesis, it was recently proposed that RNA evolved as a regulatory layer in higher organisms (Mattick, 2004) to overcome apparent scaling limitations on complexity in (protein-based) gene networks. DNA methylation also appears to be a mechanism that offers a way to limit potential connectivity of a gene to a subset of actual interaction partners. In another study (Zilberman *et al.*, 2007), experiments showed that highly methylated genes become upregulated when methylation is lost, whereas unmethylated regions often contain highly expressed genes that are regulated by specific transcription factors. Similarly, chromatin-remodeling proteins (Polo and Almouzni, 2006), such as the Polycomb group (PcG), are an important silencing mechanism that limits transcriptional regulators from binding to promoter sequence as mutations to members of PcG lead to homeotic transformations (Lewis, 1978). These results suggest that we need to be careful about assuming additional interaction partners without empirical evidence that shows a causal relationship. More generally, however, these examples suggest that the maintenance of frugal network complexity is a critical design constraint and may be an important principle for understanding biological organization and evolutionary design decisions.

## Materials and methods

### Network model

An individual is modeled as an interaction network of  $N$  transcriptional regulators encoded by an  $N \times N$  matrix  $\mathbf{W}$  (see Supplementary Figure 1), where  $w_{ij} \in \mathbf{W}$  describes the regulatory influence of gene  $j$  on gene  $i$ . Interactions ( $w_{ij} \neq 0$ )  $\in \mathbf{W}$  are assigned random variables [ $X \sim N(0, 1/3)$ ] from a discrete (rounded to  $1E-3$  significant digits) truncated ( $X \in [-1, +1]$ ) discontinuous ( $X \neq 0$ ) normal distribution. An individual's state vector  $\mathbf{s}(t) = (s_1(t), \dots, s_N(t))$  describes gene states at time  $t$  as expressed, unregulated, or repressed:  $s_i(t) = \{+1, 0, -1\}$ . Given the initial state vector  $\{\mathbf{s}_0: s_{01} = +1, \mathbf{s}_{0i} (i \neq 1) = 0\}$  and setting  $\mathbf{s}(0) = \mathbf{s}_0$ , gene states change by:

$$s_i(t+1) = \theta\left(\sum_j w_{ij}s_j(t)\right), \theta(x) = \begin{cases} +1 & x > 0 \\ -1 & x < 0 \\ 0 & x = 0 \end{cases} \quad (2)$$

Iterating (2), an individual's phenotype is its steady-state vector  $\{\hat{\mathbf{s}}: \hat{\mathbf{s}} = \mathbf{s}(t+1) = \mathbf{s}(t), t < 100\}$ . If by  $\mathbf{s}(100) \neq \mathbf{s}(99)$ , then the genotype is unviable.

### Founder population

A population is seeded with a founder genotype whose phenotype is designated as optimal:  $\hat{\mathbf{s}}^{\text{OPT}}$ . To produce a founder, random genotypes

are generated until one expresses a phenotype:  $\{\hat{\mathbf{s}}^{\text{OPT}}: \hat{\mathbf{s}}^{\text{OPT}} = \hat{\mathbf{s}}, 0 \notin \hat{\mathbf{s}}\}$ . A random genotype is generated by randomly filling a zeroed-matrix ( $\mathbf{W}$ ) with  $c_0 N^2$  non-zero values. The founder population is then seeded with  $M-1$  mutant clones of the founder (each mutated by randomly selecting, then modifying, three non-zero values).

### Fitness and selection

A Gaussian fitness-function scores fitness with:

$$f(\hat{\mathbf{s}}) = e^{-D(\hat{\mathbf{s}}, \hat{\mathbf{s}}^{\text{OPT}})/N\sigma} \quad (3)$$

Here  $D(\hat{\mathbf{s}}, \hat{\mathbf{s}}^{\text{OPT}})$  is the hamming distance between the individual and optimal phenotype. An individual's fitness gives a survival probability.

### Reproduction and mutation

By randomly mating sampled pairs, survivors populate the next generation with  $M$  offspring. Offspring inherit each row of its  $\mathbf{W}$  equiprobably from each parent. Offspring experience mutation at the rate  $\mu$ : iterating each  $w_{ij} \in \mathbf{W}$ . A network interaction is added, deleted, or modified with conditional probability:  $P(\alpha|w_{ij}=0) = \mu\alpha/N^2$ ;  $P(\phi|w_{ij} \neq 0) = \mu\phi/N^2$ ;  $P(\delta|w_{ij} \neq 0) = \mu\delta/N^2$ , where  $\phi = 1 - \delta$ .

### Measuring the average CPPI

For a matrix  $\mathbf{W}$ , the CPPI is measured as the expected deviation (scaled to unity) of the phenotype when a single non-zero element in  $\mathbf{W}$  is randomly sampled and assigned a new non-zero value. The expected deviation is estimated by randomly sampling the effects for 5000 independent perturbations. Similarly, the CPPI can be calculated as the summation of the expected cost of each interaction (CP) averaged over the number of interactions.

### Measuring GCP

For a matrix  $\mathbf{W}$ , the GCP is measured as the sum of the mean effects of perturbation on each interaction ( $w_{ij} \neq 0$ )  $\in \mathbf{W}$ , or similarly GCP can be measured as the CPPI multiplied by the number of interactions in the network.

### Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website ([www.nature.com/msb](http://www.nature.com/msb)).

## Acknowledgements

I thank Vincent Lynch and Richard Harrington for their helpful discussions; Jeremy Draghi, Alex Urban, and David Vasseur for reviewing the paper; Rimas Vaisnys for his early input into the model description; the second reviewer for his comments on the structure; and my mentor Gunter Wagner, for his sage advice and paternal rigor. This research benefited from the Yale University Life Sciences Computing Center, and NIH grant: RR19895, which funded the instrumentation.

## References

- Azevedo RB, Lohaus R, Srinivasan S, Dang KK, Burch CL (2006) Sexual reproduction select for robustness and negative epistasis in artificial gene networks. *Nature* **440**: 87–90
- Bergman A, Siegal ML (2003) Evolutionary capacitance as a general feature of complex gene networks. *Nature* **424**: 549–552
- Ciliberti S, Martin O, Wagner A (2007a) Innovation and robustness in complex regulatory gene networks. *Proc Natl Acad Sci* **104**: 13591–13596

- Ciliberti S, Martin OC, Wagner A (2007b) Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Comp Biol* **3**: 164–173
- Holland JH (1992) *Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge, Mass: MIT Press
- Huerta-Sanchez E, Durrett R (2006) Wagner's canalization model. *J Theor Pop Biol* **71**: 121–130
- Kondrashov AS (1988) Deleterious mutations and the evolution of sexual reproduction. *Nature* **33**: 435–440
- Lewis EB (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* **276**: 565–570
- MacCarthy T, Bergman A (2007a) Coevolution of robustness, epistasis, and recombination favors asexual reproduction. *Proc Natl Acad Sci* **104**: 12801–12806
- MacCarthy T, Bergman A (2007b) The limits of subfunctionalization. *BMC Evol Biol* **7**: 213
- Martin OC, Wagner A (2008) Multifunctionality and robustness trade-offs in model genetic circuits. *Biophys J* **94**: 2927–2937
- Masel J (2004) Genetic assimilation can occur in the absence of selection for the assimilating phenotype, suggesting a role for the canalization heuristic. *J Evol Biol* **17**: 1106–1110
- Mattick JS (2004) RNA regulation: a new genetics? *Nat Rev Genet* **5**: 316–323
- Noman N, Hitoshi I (2005) Reverse engineering genetic networks using evolutionary computation. *Genome Informatics* **16**: 205–214
- Oikonomou P, Cluzel P (2006) Effects of topology on network evolution. *Nat Phys* **2**: 532–536
- Polo SE, Almouzni G (2006) Chromatin assembly: a basic recipe with various flavours. *Curr Opin Genet Dev* **16**: 104–111
- Salazar-Ciudad I, Garcia-Fernandez J, Sole RV (2000) Gene networks capable of pattern formation: from induction to reaction-diffusion. *J Theor Biol* **205**: 587–603
- Siegal ML, Bergman A (2002) Waddington's canalization revisited: developmental stability and evolution. *Proc Natl Acad Sci* **99**: 10528–10532
- Siegal ML, Promislow D, Bergman A (2006) Functional and evolutionary inference in gene networks: does topology matter? *Genetica* **129**: 83–103
- Solé RV, Fernandez P, Kauffman SA (2003) Adaptive walks in a gene network model of morphogenesis: insights into the Cambrian explosion. *Int J Dev Biol* **47**: 685–693
- von Dassow G, Meir E, Munro EM, Odell GM (2000) The segment polarity network is a robust development module. *Nature* **406**: 188–192
- Wagner A (1996) Does evolutionary plasticity evolve? *Evolution* **50**: 1008–1023
- Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**: 61–69



*Molecular Systems Biology* is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*.

This article is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Licence.