Review

Modeling Genetic Networks and Their Evolution: A Complex Dynamical Systems Perspective

Stefan Bornholdt

Institut für Theoretische Physik, Universität Kiel, Leibnizstrasse 15, D-24098 Kiel, Germany

After finishing the sequence of the human genome, a functional understanding of genome dynamics is the next major step on the agenda of the biosciences. New approaches, such as microarray techniques, and new methods of bioinformatics provide powerful tools aiming in this direction. In the last few years, important parts of genome organization and dynamics in a number of model organisms have been determined. However, an integrated view of gene regulation on a genomic scale is still lacking. Here, genome function is discussed from a complex dynamical systems perspective: which dynamical properties can a large genomic system exhibit in principle, given the local mechanisms governing the small subsystems that we know today? Models of artificial genetic networks are used to explore dynamical principles and possible emergent dynamical phenomena in networks of genetic switches. One observes evolution of robustness and dynamical self-organization in large networks of artificial regulators that are based on the dynamic mechanism of transcriptional regulators as observed in biological gene regulation. Possible biological observables and ways of experimental testing of global phenomena in genome function and dynamics are discussed. Models of artificial genetic networks provide a tool to address questions in genome dynamics and their evolution and allow simulation studies in evolutionary genomics.

Key words: Complex systems / Evolutionary genomics / Functional genomics / Genetic networks / Genome-wide modeling / Simulation studies / Self-organization.

Introduction

While the completed sequence of the human genome as today's largely unreadable Rosetta stone awaits deciphering, the first milestones have been passed in the adventure of deciphering genome function. Examples are the development of microarray techniques that currently revolutionize gene expression studies on up to genome-wide scales, as well as advanced data analysis tools to reconstruct gene regulatory interactions. Several interesting model systems are currently being studied where first genome-wide expression analyses provide a complementary view on gene functioning, well beyond the one-gene-one-protein perspective. Furthermore, detailed studies of separable genetic subcircuits as functional modules of genomes are successfully performed with a number of model organisms.

However, a full understanding of genome dynamics on a larger scale is not as easily available and will probably require a much bigger effort. Not only experimentalists, but also the theoretical sciences feel challenged by this problem. They study possible approaches to an understanding of genome dynamics and how a theorist can contribute to the toolbox of molecular biology and bioinformatics. One such approach is a systems scale view of the genome as a complex interacting system of many components. Indeed, mathematical and physical sciences have found ways to approach complex dynamic systems in various branches of science, and one may ask whether such approaches could be applicable to the genome and the new challenges of data-driven branches of the biosciences. For example, comparing the basic mechanism of transcriptional regulators to that of simple switches makes approaches applicable to basic questions in gene regulation that study complex dynamic systems and use tools from theoretical physics. An interesting question is what could be in principle the dynamics of large systems of interconnected (genetic) switches? While currently experimentally inaccessible in the regulatory circuit of a genome, such a question can well be answered in theoretical model systems of many switches. One possible approach to such questions by modeling artificial genetic networks will be sketched in this paper. One question studied in such models directly addresses the dynamics of networks of regulatory genes, as observed in cell control and differentiation. A second complex of questions addresses evolutionary genomics and how gene regulation interacts with biological evolution, as for example seen in speciation in the face of a strong requirement of stability of genetic networks.

In this article, two such studies will be described and their implications for an understanding of genome-wide gene regulation is discussed. First, a brief review of current developments in genome-wide expression studies is given, followed by a brief overview of current bioinformatics approaches to genetic network reconstruction. Then theoretical problems connected to the dynamics of genome-wide gene regulation are discussed

before embarking on simulation studies addressing two of these questions.

Genome-Wide Gene Expression Data

Let us briefly review some of the data of genome-wide expression studies that provide information about the building blocks of global genome organization. For a number of model organisms, microarray techniques provide genome-wide expression analyses, for example in yeast, where the expression pattern of the entire genome can be monitored over time (DeRisi *et al.*, 1997). A comprehensive analysis in yeast (*Saccharomyces cerevisiae*) shows that about 250 regulators interact with the transcription and expression of over 6000 genes (Holstege *et al.*, 1998). By gene deletions or *via* temperature sensitive genes one obtains differential snapshots that provide functional information about correlated activity of genes. In the nematode *C. elegans*, about two thirds of the genes can be assayed (Kim, 2000) and even in *Drosophila* expression analyses reaches a near genome-wide scale with about half the genome (White *et al.*, 1999).

One observation generally made in such studies is the highly coordinated expression of genes which can often be classified in groups of co-expressed genes that share a common (and often complex) temporal expression pattern (synexpression groups). Moreover, such groups of genes can often be tied to a specific function as, for example, in the cell cycle, metabolism, protein biosynthesis or others (Niehrs and Pollet, 1999). The observed abundant modular organization of genetic networks, which could be said to act as subroutines of biological programs, is a remarkable feature of genome organization.

While reminiscent of the prokaryote operon, co-regulated genes in these groups often are located on different chromosomes, which is another interesting observation concerning genetic network organization. Other than the *cis*-regulated operon, groups of co-regulated genes in eukaryotes are more likely to be regulated by one or more common *trans*-regulators. Indeed, such common regulators can be traced by genome-wide expression analysis. A powerful technique is a combination with upstream DNA motif search for all member genes of a group to identify known as well as unknown regulatory elements of this group (Tavazoie *et al.*, 1999). Also *vice versa*, for a given regulator, all genes regulated by this factor can be found by genome-wide analysis (Ren *et al.*, 2000). Beyond simple wiring, such analyses are able to reveal hierarchies in gene regulation as in the case of high level regulators that control large parts of gene expression, as for example a regulator that controls large parts of a bacterial cell cycle (Laub *et al.*, 2000). From the perspective of theory, these are exactly the basic steps needed for the inference for causal relationships between genes (whereas plain expression analysis always can only provide correlation). However, simple combinatorial arguments set strong limits for the size of sub-networks that can be reconstructed by this method.

A further remarkable observation in genome-wide analysis is the recent estimate of the total number of genes in humans of 35 000 (Ewing and Green, 2000). Since this is only about half the number than has been assumed earlier and not even twice the number of genes found in the nematode *C. elegans*, the larger physiological complexity of a human cannot be attributed to additional genes to the same extent as has been common view until recently. Instead, the role of regulatory combinatorics and the diversification of genetic networks seems to play a much larger role than thought before. This is one reason for a further study of possible dynamic processes in large regulatory networks.

Reconstruction of Global Genetic Network Properties

Bioinformatics algorithms for gene expression modeling face considerable difficulties, from the high quantitative error of expression experiments to an insufficient number of data points when aimed at genetic network reconstruction. For genetic subnetworks that are mostly modular and only loosely connected to the rest of the genome, modeling works quite well, in particular clustering of expression data into groups of co-expressed genes (Wen *et al.*, 1998; Basset *et al.*, 1999) as also the above examples show. In networks where data are not as well separated, however, clustering often is ambiguous. Also, clustering relies on strictly linear gene-gene interactions (Bittner *et al.*, 1999).

Beyond clustering co-expressed genes from array data, we finally are faced with the underlying complexity of genome-wide function. In a simple toy model at least, one finds that this problem may finally not persist in the experiment: a simple estimate of the number of required experiments for a full reconstruction of a network shows that only about $K \log(N)$ microarray experiments would be necessary for an approximate reconstruction of a full network, where K is the average number of regulatory genes that affect a given gene and N is the overall number of genes (Hertz, 1998). While this is clearly more than currently available datasets can offer, experiments of this size are not inconceivable in the future. How this problem scales for real genetic networks and in the face of noisy data is, however, quite open today.

A conceptual problem in the reconstruction of genetic networks from raw expression data that remains unsolved is the trivial fact that measuring correlation in general is not sufficient to infer causality between genes. Here, a combination of algorithms that closely interact with experimental data could obtain causal information (D'haeseleer *et al.*, 2000).

As predictive genetic network models will be out of reach for quite some time it will be worthwhile to ask some more basic questions about what principle dynamical properties such model networks can exhibit. For such an approach we will use types of networks that are

also used in the above reconstruction approaches and their basic assumptions about transcriptional regulation.

A Systems Science View on Complex Biological Networks

What can we learn from other complex dynamical systems about gene regulation in the genome? One interesting dynamical property often observed in complex interacting networks is their ability to self-organize to an emergent dynamical state that is not easily derived from their local interaction and binding constants. One example is the observed robustness in bacterial chemotaxis which is maintained by a network of interacting proteins (Alon *et al.*, 1999). The achieved adaptation is maintained at a very precise level, which proves to be largely independent of individual protein concentrations. Biochemical reaction networks fulfill all requirements for complex emergent properties as, for example, information storage and learning (Bhalla and Iyengar, 1999).

Let us pose the question of how such emergent phenomena stabilize in complex networks and how such networks themselves emerge from evolution in a more general setting. What can we learn from the general theory of complex dynamical networks (Strogatz, 2001)? Complexity in network structures can be manifold: the nodes of the network may be different and diverse, as also the interaction links between the nodes may be. Structural complexity of a network shows up in a complex wiring diagram. Furthermore, dynamical complexity may show up in the dynamics of the network, for example when nodes interact in complicated ways. But also the network may evolve over time (as genetic networks do on evolutionary time scales) and if this is coupled to the network dynamics itself, a quite complicated overall complex network results. This is about the level of complexity of a biological genetic network.

Tools for treating such complex dynamical networks have been developed in the fields of graph theory and statistical physics with a tradition of modeling complex neural networks. Several remarkable global properties of networks are known in this context. In some types of networks phase transitions between regimes of ordered and disordered dynamics have been described that may be relevant to biological systems (Kauffman, 1993). Another theoretical concept is percolation theory which is useful for describing how information spreads over a network (Stauffer and Aharony, 1995). An important question is the robustness of the dynamics of networks against disturbing their dynamics or architecture.

In the following I will discuss how specific properties of genetic networks could result from biological evolution. An interesting set of questions is how the complex network of genetic interactions influences the dynamics of macroevolution and whether it generates boundary conditions on evolution and speciation. As we know from the evolution of homeobox genes, basic features of high lev-

el genetic networks evolve at an astonishingly slow speed (Pennisi and Roush, 1997; Knoll and Caroll, 1999). Also, their structure allows single gene mutations to introduce major innovations, and it exhibits a modular structure, presumably to preserve the major part of the genome against mutations. A long-term stability of a gene cluster is observed in the Hox complex which exhibits an extremely tight binding over evolutionary times (Patel and Prince, 2000). In general, groups of co-regulated genes, as observed in many expression studies, also pose questions on the evolutionary scale: what is the selective pressure that promotes this type of genetic organization? Again, the modular organization of function may be advantageous in the face of evolution and major functional innovations may be linked to changes in a single gene, as is discussed for the expression of digits at the same time in hand and foot (Niehrs and Pollet, 1999). These and other issues are possible questions to address when modeling the evolution of artificial genetic networks.

For this purpose, let us discuss models of artificial genetic networks and simulation studies of their evolution. One of the first simulation approaches from the statistical physics perspective has been formulated by Wagner (1996) who modeled a network of transcriptional elements in a spin glass or neural network-type model. The model genes mutually regulate each other and produce stable gene expression patterns. In this model, the stability of the epigenetic system to mutations is studied. For this purpose, populations of small model networks (N=4..10) are evolved in the presence of a fitness function, that selects for a maximal overlap with a given master expression pattern, serving as a given optimal expression pattern. The expression patterns for simplicity are chosen to be fixed points of the dynamics (while other models consider limit cycle attractors representing dynamical expression patterns). One advantage of the principal approach of such models ist that one can study general situations in a statistical sense, *i. e.* consider ensembles of systems that give an estimate of what an average behavior of systems would be, without knowing their specificities. Here, one considers ensembles of random networks, yielding the average behavior of networks sharing the same general features. This makes the results more applicable to real natural systems. After an evolution process of a few hundred generations, this model finds enhanced mutational stability of the evolved networks in the sense that the expression patterns of the networks are more robust against mutations and less frequently destroyed by a mutation as compared with random networks. Partly, this is accounted for by shorter expression patterns, but one also observes an additional inherent stability for evolved networks when compared to random nets with the same length of expression patterns. This phenomenon is termed epigenetic neutrality as opposed to molecular neutrality denoting molecular stability against sequence mutations. One interesting aspect of this phenomenon is its possible influence on the

maximum rate of evolution. A second interpretation of this work is that the increased stability of the regulatory network may buffer the phenotype against mutations. In this sense, epigenetic stability is enhanced: a mutation of the gene regulatory circuit will affect its expression pattern less often in the evolved networks than in random networks. Of course, any such event would be most likely deleterious or fatal if it occurs during development of a such regulated organism.

While this work is based on the definition of an artificial fitness, one may step back and ask what the corresponding force in a biological system would be. Moreover, macroevolution as a whole may not operate on a fitness landscape at all. Clearly, fitness is a common concept in evolution (Wright, 1982), and often evolution is simply viewed as hill climbing and with jumps between fitness maxima (Newman, 1985; Lande, 1985). Nevertheless, fitness landscapes implicitly assume that fitness obeys a well-defined metric in genomic space. This is only the case if single point mutations were a driving force and obviously would require the absence of gene regulation. However, significant genome rearrangements are already observed in real-time evolution experiments of *E. coli* cultures (Papadopoulos *et al.,* 1999). Genomic rearrangements such as deletions or insertions destroy the simple metric generated by one-point mutations, which usually underlies the intuition of evolution on landscapes. As a consequence the combinatorial distance for moving from one genome to another may very well be different from the distance of the opposite move as, for example, a deletion is easily made, but hard to recover once lost. Thus, although fitness landscapes have a well defined meaning for the small scale adjustments associated to fine-tuning of binding constants, it is an unjustified concept for evolutionary changes on the macroevolutionary scale of speciation events.

Modeling Genetic Network Evolution Without Fitness

A computer study of artificial neutral genetic network evolution without any explicit definition of a fitness function which explores further this viewpoint has first been published in (Bornholdt and Sneppen, 1998, 2000). It will be recapitulated in this section. An observation that challenges the role often ascribed to fitness in evolution is the fact that one often observes different phenotypes for the same genotype, as enabled by gene regulation and observed in such diverse examples as cell differentiation, metamorphosis, and other epigenetic phenomena. An important non-trivial mechanism for evolution may thus be the exposure of the same species to different environments. The species then faces a variable selection criterion, with the consequence that what is phenotypically neutral at some instant may not be phenotypically neutral at later instants. Thus, in contrast to the molecular neutrality where many RNA genotypes have the same phenotype (Schuster, 1997), in genetic network neutrality more than one phenotype for each genotype may occur.

In the following, a class of model systems is studied that exhibits epigenetics as a simple model for transcriptional regulation. It is represented by logical networks, where nodes in the network take values on or off, as a function of the output of specified other nodes. In terms of these models it is natural to define genotypes in form of the topology and rules of the nodes in the network. The phenotypes are similarly associated to the dynamical expression patterns of the network. As a prerequisite a model for evolution should fulfill the requirement of robustness. Robustness is defined as the ability to function in spite of substantial change in components (Savageau, 1971; Hartwell, 1997; Alon *et al.*, 1999; Little *et al.*, 1999). Robustness is an important ingredient in simple molecular networks and probably also an important feature of gene regulation on both, small and large scale. In the framework of an evolutionary model based on logical networks, robustness is implemented by requiring that mutations of the regulatory network do not change expression patterns.

Network types that exhibit epigenetics are Boolean networks (Kauffman, 1969), and a subset of those are the threshold networks (Kürten, 1988a, b). In these networks each node takes on one of two discrete values, $\sigma_i = \pm 1$, that at each time step is a function of the value of some fixed set of other nodes. The links that provide input to node i are denoted by $\{w_{ij}\}$ with $w_{ij} = \pm 1$. A crucial structural parameter of the network is its connectivity K, defined as the average number of incoming (non-zero) weights per node. The updating rule for the dynamics on the network is defined as follows: For the threshold network case it is additive:

$$\sigma_i = \quad 1 \;\; if \sum_{j \in \{w_i\}} w_{ij}\sigma_j \geq 0 \tag{1}$$

$$\sigma_i = -1 \;\; if \sum_{j \in \{w_i\}} w_{ij}\sigma_j < 0 \tag{2}$$

The threshold networks are well known as a type of neural networks, where a certain number of input firings are necessary to induce firing in a given neuron (Kürten, 1988a, b). Boolean networks are mostly discussed in connection with genetic networks, as the specificity of protein binding in principle enables the implementation of more detailed logical functions. On the other hand, threshold networks to a good approximation represent the basic principle of transcriptional regulation (Wagner, 1996). The basic property of logical networks is a dynamics of the state vector $\{\sigma_i\}$ characterized by transients that lead to subsequent attractors, the periodic activity pattern to which the network dynamics converges. The attractor length depends on the topology of the network. Below a critical connectivity $K_c \sim 2$ (Derrida and Pomeau, 1986; Kauffman, 1993) the network decouples into many disconnected regions, *i. e.*, the corresponding genome expression would become modular, with essentially independent gene activities. Above K_c any local damage will initiate an avalanche of activity that may propagate

throughout most of the system. For any K above K_c the attractor period diverges exponentially with respect to the number of nodes N, and in some interval above K_c the period length in fact also increases nearly exponentially with connectivity K (Bastola and Parisi, 1996). Note that here the critical connectivity (or coordination number) equals 2, compared to unity in usual random graphs (Erdös and Renyi, 1960; Bollobas, 1985), due to the Boolean logic. Criticality means that a change at a node in the network spreads marginally throughout the network. This picture is particularly simple for Boolean networks where any activity change of a node has the probability 1/2 to propagate along any link for random Boolean rules, so that an average of 2 links have to leave each node to create the critical state. For threshold networks similar arguments apply.

The evolution of the network topology is defined as a change in the wiring $\{w_{ij}\} \rightarrow \{w'_{ij}\}$ that takes place on a much slower time scale than the $\{\sigma_i\}$ updating. The evolution of such networks represents the extended degree of genetic network engineering that seems to be needed to account for the large differences in the structure of species genomes (Shapiro, 1998), given the slow and steady speed of single protein evolution (Kimura, 1983). The model will extend neutral evolution on the molecular scale to neutral evolution on the regulatory level, and demonstrate that neutrality in itself enforces constraints on the evolved graphs.

First it has been proposed to evolve Boolean networks with the sole constraint of continuity in expression pattern (Bornholdt and Sneppen, 1998). Later this model has been simplified to transcriptional regulators combined with a simple test of damage spreading (Bornholdt and Sneppen, 2000): the model evolves a new single network from an old network by accepting rewiring mutations with a rate determined by expression overlap. This is a minimal constraint scenario with no outside fitness imposed. Furthermore, the model tends to select for networks which have high overlap with neighbor mutant networks, thus securing robustness. The model is defined as follows: consider a threshold network with N nodes. To each of these a logical variable $\sigma_i = -1$ or $+1$ is assigned. The states $\{\sigma_i\}$ of the N nodes are simultaneously updated according to (1, 2) where the links w_{ij} are specified by a matrix. The entry value of the connectivity matrix w_{ij} may take the values -1 and $+1$ in case of a link between i and j, and the value 0 if i is not connected to j. The system that is evolved is the set of couplings w_{ij} in a single network. One evolutionary time step of the network is:

(1) create a daughter network by (a) adding, (b) removing, or (c) adding and removing a weight in the coupling matrix w_{ij} at random, each option occurring with probability $p = 1/3$. This means turning a $w_{ij} = 0$ to a randomly chosen ± 1 or *vice versa*.

(2) Select a random input state $\{\sigma_i\}$. Iterate simultaneously both the mother and the daughter system from this state until they either have reached and completed the same attractor cycle, or until a time where $\{\sigma_i\}$ differs be-

tween the two networks. In case their dynamics is identical then replace the mother with the daughter network. In case their dynamics differs, keep the mother network.

Thus, the dynamics looks for mutations which are phenotypically silent, *i. e.*, they are neutrally inherited under at least some external condition. Note that adding a link involves selecting a new w_{ij}, thus changing the rule on the same time scale as the network connectivity. Iterating these steps represents an evolution which proceeds by checking overlap in expression pattern between networks. If there are many states $\{\sigma_i\}$ that give the same expression of the two networks, then transitions between them are fast. On the other hand, if there are only very few states $\{\sigma_i\}$ which result in the same expression for the two networks, then the transition rate from one network to the other is small. If this is true for all its neighbors then the evolutionary process will be hugely slowed down. Interestingly, other than in existing concepts of selective neutrality (Schuster 1997), these transition rates are not constant in this model of regulatory neutrality. In particular, they are a function of the evolving connectivity K of the network instead.

In Figure 1 the basic elements of the model are shown: only mutations of the regulatory network are accepted that do not change the present expression pattern. In the resulting evolution one observes that in particular for highly interconnected networks the system may stay a long time at a particular network before an allowed mutation leads to punctuations of the stasis. This waiting time until a successful mutation occurs is a characteristic quantity of the evolutionary process. The overall distribution of waiting times is $\sim 1/t^{2\pm0.2}$ on the evolutionary time scale of the model where t counts the complete iteration cycles of the model. The wide variety of time scales implied by the $1/t^2$ distribution reflects the different time scales that are associated with networks of different connectivity K. Thus, any particular network will have a characteristic time scale with exponentially distributed waiting time. The $1/t^2$ distribution originates from integration over this broad range of time scales, reflecting that the probability of accepting a mutation decreases exponentially with K, whereas the probability per attempt to add a specific link equals the probability to remove it again.

One important feature of the evolution is the structure of the evolved networks, which can be quantified by the average length of expression patterns (attractors) for the generated networks. This is shown in Figure 2, where they are compared with attractor lengths for random networks at the same connectivity. One observes that the evolved networks have much shorter attractors than the random ones, thus the evolution favors simplicity of expression. To examine further the expression behavior of the networks consider the size of frozen components as introduced by Kauffman (1990) for Boolean networks. A frozen component is the set of nodes connected to a given attractor that does not change at any time when you iterate along the attractor, *i. e.*, a frozen component repre-

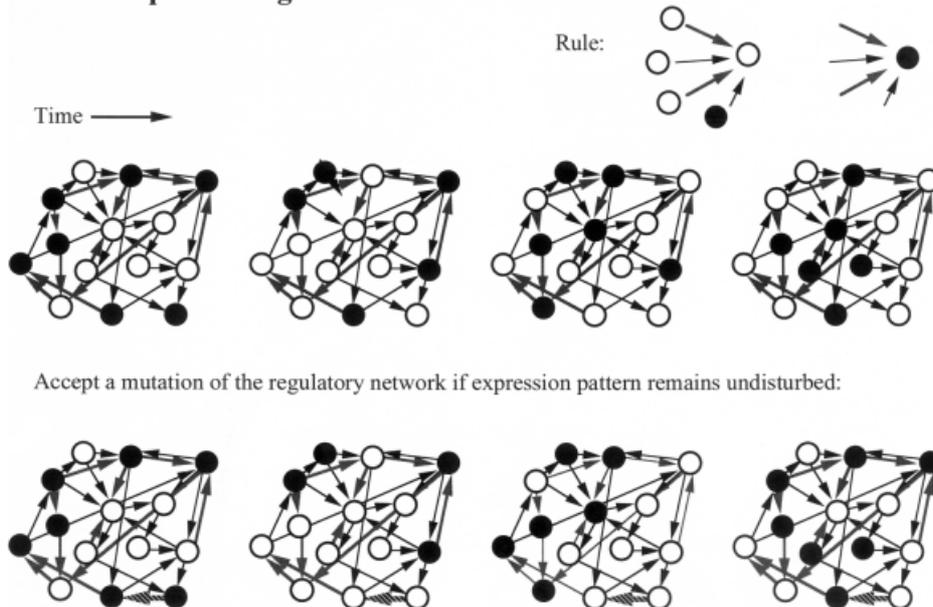## Transcriptional regulation as networks of switches

Fig. 1 A Schematic View of the Evolution Model as Defined by Bornholdt and Sneppen (2000).
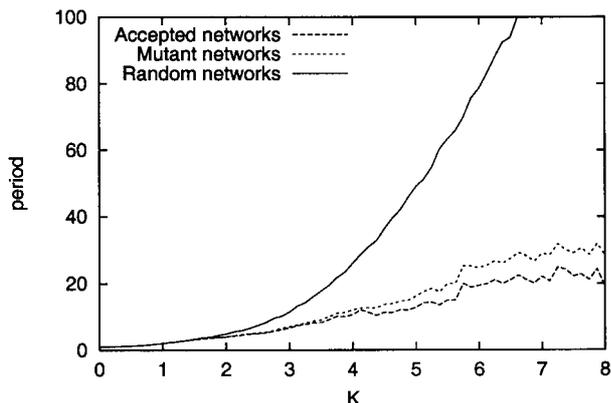Only mutations of the regulatory network are accepted that do not change the current expression pattern.

Fig. 2 Average Length of Expression Patterns (Periodic Attractors) for Evolved and for Random Networks.
Also the periods of the unsuccessful mutations in the presence of newly chosen random initial conditions are shown, demonstrating that selection of networks is indeed operating on network structure and the specific input configuration in the event of selection does not play a major role.

Fig. 3 Average Size of Frozen Components as a Function of Connectivity for Evolved and Random Networks.
The frozen component is the set of all genes that are not part of the current expression pattern. One observes that the robustness constraint in evolution favors a larger frozen component.

sents the fraction of genes which are silent under a given attractor/initial conditions. In Figure 3 it can be seen that the frozen component for the evolved network typically involves half the system, and thus is much larger than the typical frozen component associated to attractors of randomly generated threshold networks. Also random one-mutant neighbors of the selected networks have huge frozen components. Finally, the active part of the network and the complexity of its expression pattern is an interesting observation. As a quite large fraction of the nodes may belong to the frozen component of the network, the remaining active part of the nodes may behave different-
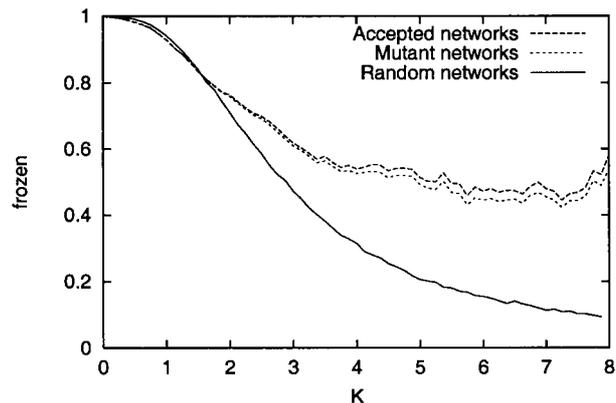
ly from the average dynamics of the whole network. One possible measure is the number of times each non-frozen node switches its state during the dynamical attractor. In Figure 4 this quantity is shown for random networks as well as evolved networks. One observes that the active part of the evolved networks exhibits a much simpler expression pattern than that of a random network of comparable connectivity.

To summarize, the results of this genetic network evolution model, implementing robustness as an evolutionary criterion, has observable consequences for both, the temporal evolution pattern, and for confining possible genetic network architectures to the ones with simple expression patterns.
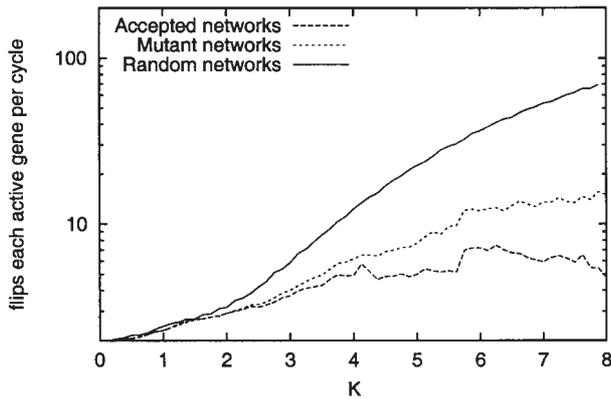
Fig. 4 Average Number of Flips per Gene in the Non-Frozen Part of the Network as a Function of Connectivity for Evolved and for Random Networks.

The evolved networks show a reduced activity in the non-frozen genes resulting in simple expression patterns as compared to those of random networks of same connectivities. Note that the number counts off-on and on-off transitions of the genes as separate events.

Self-Organization of Genetic Networks from Local Adaptations

Let us now turn to a second model of evolving genetic networks where instead of observing global expression patterns we concentrate on local functional changes of single genes and their influence on global network architecture and dynamics. This model has been first published by Bornholdt and Rohlf (2000) and will be summarized in the following. Again, logical threshold networks will be considered as a model for transcriptional regulation. In these networks one in general observes a phase transition at a critical average connectivity K_c with lengths of transients and attractors (limit cycles) diverging exponentially with system size for an average connectivity larger than K_c. Combinatorial as well as numerical methods provide a quite detailed picture about their dynamical properties and correspondence with Boolean Networks (Derrida and Pomeau, 1986; Kürten, 1988a, b; Bastolla and Parisi, 1996; 1998). While basic dynamical properties of interaction networks with fixed architecture have been studied with such models, the origin of specific structural properties of networks in natural systems is often unknown. For genetic networks Kaufmann postulated that gene regulatory networks may exhibit properties of dynamic networks near criticality (Kauffman, 1969, 1993). However, this postulate does not provide a mechanism able to generate an average connectivity near the critical point. An interesting question is whether connectivity may be driven toward a critical point by some dynamic mechanism. In the following such an approach will be sketched in a setting of an explicit evolution of the connectivity of networks.

Let us consider again a network of N randomly interconnected binary elements as defined above. The dy-

namics of the network states is again obtained by iterating the threshold rule starting from a random initial condition, eventually reaching a periodic attractor (limit cycle or fixed point).

Then the following local rewiring rule is applied to a randomly selected node i of the network: if node i does not change its state during the attractor, it receives a new non-zero link c_{ij} from a random node j. If it changes its state at least once during the attractor, it loses one of its non-zero links c_{ij}. Iterating this process leads to a self-organization of the average connectivity of the network. To be specific, define the average activity $A(i)$ of a site i as

$$A(i) = \frac{1}{T_2 - T_1} \sum_{t=T_1}^{T_2} \sigma_i(t) \qquad (3)$$

where the sum is taken over the dynamical attractor of the network defined by T_1 and T_2. The algorithm then iterates the following steps:

(1) Choose a random network with an average connectivity K_{ini}.

(2) Choose a random initial state vector $\vec{\sigma}(0) = (\sigma_1(0),...,\sigma_N(0))$.

(3) Calculate the new system states $\vec{\sigma}(t)$, $t = 1,...,T$ according to eq. (2), using parallel update of the N sites.

(4) Once a previous state reappears (a dynamical attractor is reached) or otherwise after T_{max} updates the simulation is stopped. Then change the topology of the network according to the following local rewiring rule:

(5) A site i is chosen at random and its average activity $A(i)$ is determined.

(6) If $|A(i)|=1$, i receives a new link c_{ij} from a site j selected at random, choosing $c_{ij}=+1$ or -1 with equal probability. If $|A(i)|<1$, one of the existing non-zero links of site i is set to zero.

(7) Go to step number 2 and iterate.

The basic mechanism of the model is further shown in Figure 5. The typical dynamics arising from the model as defined above is shown in Figure 6 for a system of size $N = 1024$. Independent of the initial connectivity, the system evolves toward a statistically stationary state with an average connectivity $K_{ev}(N = 1024) = 2.55 \pm 0.04$. In the large system size limit $N \rightarrow \infty$ the networks evolve to a value close to the critical connectivity of the network $K_c \approx 2$.

The self-organization toward criticality observed in this model is different from currently known mechanisms exhibiting the general phenomenon of self-organized criticality (SOC) (Bak et al., 1987; Bak and Sneppen, 1993). This model introduces a new type of mechanism by which a system self-organizes toward criticality, here $K \rightarrow K_c$. In particular, it exhibits considerable robustness against noise in the system. The main mechanism here is based on a topological phase transition in dynamical networks. To see this consider the statistical properties of the average activity $A(i)$ of a site i for a random network. It is closely related to the frozen component $C(K)$ of the network, defined as the fraction of nodes that do not change their state along the attractor. The average activity $A(i)$ of a frozen site i thus obeys $|A(i)|=1$. In the limit of large N, $C(K)$

Self-organized critical network dynamics

Time——▶

Dynamical rule:
1. Track a single gene

2. Inactive genes acquire a function (or else vanish), too active genes lose some functionality:
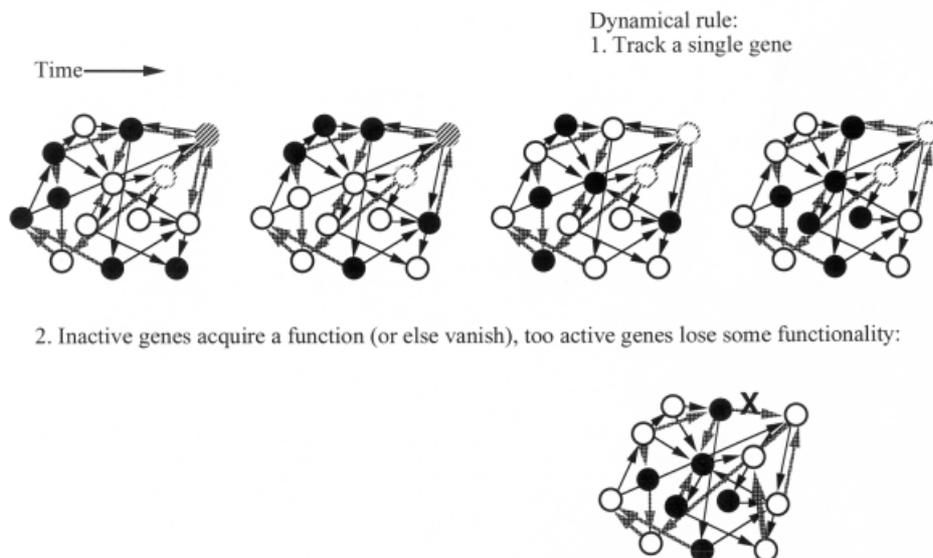

Fig. 5   A Schematic View of the Evolution Model as Defined by Bornholdt and Rohlf (2000).
Interaction strengths between regulatory genes are mutated locally, observing the constraint that a gene never loses all his functions, but may lose some of its functionality if it is dynamically very active. This selective pressure, active at the one gene level, leads to a genome-wide self-organization of the network to a dynamical regime between order and chaos.
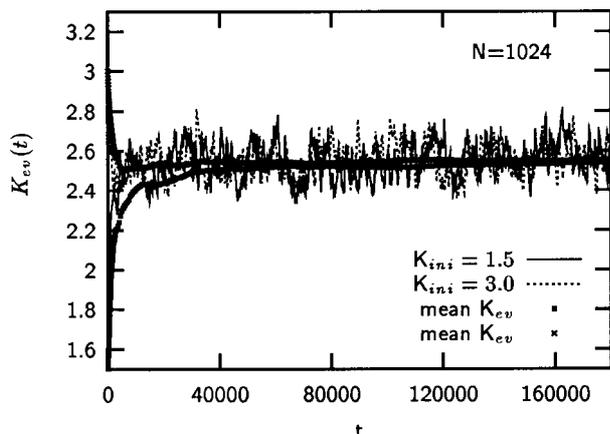

Fig. 6   Evolution of the Average Connectivity for $N = 1024$ and Two Different Initial Connectivities (K_{ini} =1.5 and K_{ini} =3.0). Independent of initial conditions the networks evolve to an average connectivity $K_{ev} = 2.55$. The plot shows the time series and the corresponding cumulative means for K_{ev}. The evolutionary time t is discrete, each time step representing a dynamical run on the evolved topology.
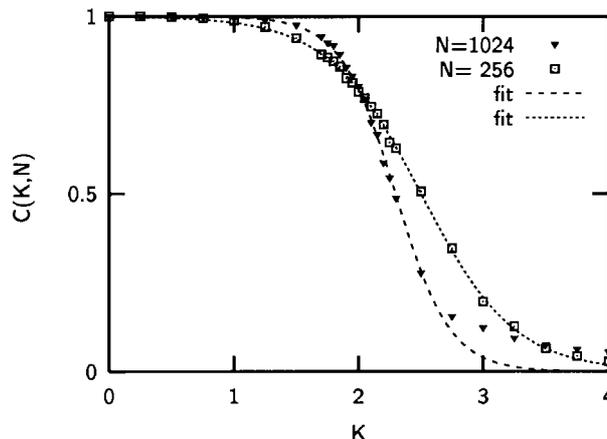
Fig. 7   The Frozen Component $C(K,N)$ of Random Threshold Networks as a Function of the Networks' Average Connectivities K.
One observes a transition near the critical connectivity of the network.

$$C(K,N) = \frac{1}{2} \{1 + \tanh [-\alpha(N) \cdot (K - K_0(N))]\}. \qquad (4)$$

undergoes a transition at K_c vanishing for larger K. With respect to the average activity of a node, $C(K)$ equals the probability that a random site i in the network has $|A(i)|=1$. Note that this is the quantity which is checked stochastically by the local update rule in the above algorithm. The frozen component $C(K,N)$ is shown for random networks of two different system sizes N in Figure 7. One finds that $C(K,N)$ can be approximated by

This describes the transition of $C(K,N)$ at an average connectivity $K_0(N)$ which depends only on the system size N. One finds that in the thermodynamic limit $N \rightarrow \infty$ the transition from the frozen to the chaotic phase becomes a sharp transition near the critical connectivity K_c. These considerations apply well to the evolving networks in the rewiring algorithm.

In addition to the rewiring algorithm above a number of

different versions of the model work as well. Including the transient in the measurement of the average activity $A(i)$ results in a similar overall behavior (where we allowed a few time steps for the transient to decouple from initial conditions). Another version succeeds using the correlation between two sites instead of $A(i)$ as a mutation criterion. In addition, this version was further changed allowing different locations of mutated links, both, between the tested sites or just at one of the nodes. All these different realizations exhibit the same basic behavior as found for the model above. Thus, the mechanism exhibits considerable robustness, a prerequisite for applicability to biological systems.

Genetic Network Models and Experiment

Let us briefly discuss results and implications of the evolution scenarios for artificial transcriptional regulation networks which have been studied in the two previous sections. What do we learn from such simulation approaches to genetic network evolution and dynamics? Let us review possible implications for biological systems.

For the neutral evolution scenario, a link to macroevolution can be drawn as the intermittent evolution of the networks is reminiscent of punctuated equilibrium as observed for species in the fossil record (Gould and Eldredge, 1993).Quantitatively, the $1/t^2$ distribution of lifetimes for single networks that one finds for this model, as well as for the earlier version (Bornholdt and Sneppen, 1998), compares well with similar scaling observed for the statistics of birth and death of individual species in the evolutionary record (Bak and Sneppen, 1993). In fact, the analogy can even be fine-grained into a sum of characteristic lifetimes, each associated to a given structural feature of the networks (Bornholdt and Sneppen, 1998). A similar decomposition is known from the fossil record (VanValen, 1973), where groups of related species display Poisson-distributed lifetimes and, therefore, similar evolutionary stability.

Testing the models at the molecular level of gene regulation can be based either on direct probing of genetic networks, but also on evolution experiments of fast-lived organisms such as *E. coli* (Papadopoulos *et al.*, 1999). Information on the overall organization of these genetic networks is obtained from correlated gene knock-out experiments. A quantitative estimate for the overall degree of connectivity in the genome can be deduced from Elena and Lenski's experiments (1999) on double mutants, which demonstrated that about 30–60% of these (dependent on interpretation) change their fitness in a cooperative manner. In terms of the artificial network models, one should expect a coupled genetic expression for about half of the of pairs of genes. Although the evolved networks can give such correlations for current connectivity estimates, the uncertainty is still so large that random networks also are in accordance with data. Further

one should keep in mind that the *E. coli* genome is large and not well represented by threshold dynamics of all nodes, and also that only between 45 and 178 of the *E. coli*'s 4290 genes are likely to mediate regulatory functions (Blattner, 2000). Thus, most of the detected gene-gene correlations presumably involve genes which are not even regulatory, but instead metabolic and their effect on each other more indirect than in the case of the regulatory ones. One would obtain stronger elements of both, coupling and correlation, if one specialized on regulatory genes. Thus one may wish to perform experiments where one- and two-point mutations are performed in regulatory genes only. A more direct test of the hypothesis of robustness in form of damage control as a selection criterion may be obtained from careful analysis of the evolution of gene regulation in evolving *E. coli* cultures.

A further recent experimental approach is the study of the divergence of duplicate genes and the divergence of their expression patterns. In a study in yeast (Wagner, 2000a) it was observed that the expression patterns of duplicate genes diverge at speeds almost uncorrelated to the divergence of the original sequence, pointing to a high flexibility on the genetic network level. Again, for the computer experiments discussed here only coupled knock-out experiments would be conclusive, which would be particularly interesting in duplicate genes.

Another interesting experimental observation is the simplicity of biological expression patterns. For example as observed in yeast many genes are only active one or two times during the expression cycle (Cho *et al.*, 1998), thus switching from off to on or on to off occurs for each gene in this system only a few times during expression. For random dynamic networks of comparable size one would expect a much higher activity. Thus surprisingly simple expression patterns are observed in biological gene regulatory circuits. This bears resemblance with the first model's observation where simplicity of expression patterns emerges as a result of a the evolutionary constraint of robustness.

A common observation of the models discussed above is the emergence of networks that are mutationally robust compared to random networks. A similar observation is made experimentally in yeast where the robustness of the gene regulation networks against single gene mutations has been tested (Wagner, 2000b). A main observation is that single gene mutations are often phenotypically silent, possibly due to a buffering of the intact gene regulation circuit for this single error. Wagner's study seems to indicate that quite unrelated genes are major agents in this buffering, rather than quasi-redundant copies of the mutated genes in the form of closely related genes. As an effect, this might be an evolved response of the global genetic network to stabilizing selection.

A further key observation is the estimated average connectivity K of $2 \rightarrow 3$ in the *E. coli* genome (Thieffry *et al.*, 1998). The second model of genetic network evolution by local adaptations demonstrates how such an in-

termediate connectivity of a regulatory network may emerge by self-organization. With respect to genetic networks one may discuss whether biological evolution exerts selection pressure on the single gene level, that results in a selection rule similar to that model. Namely, for a frozen regulatory gene which is practically non-functional to obtain a new function (obtain a new link), as well as for a highly active gene to reduce functionality (remove a link). It is interesting to note that the robust self-organizing algorithm described here provides a mechanism that in principle predicts a value in the observed range.

Summary and Outlook

In this article, approaches to an understanding of global genome organization have been reviewed, including the current state of experimental approaches and an emphasis on computer simulations of evolving artificial genetic networks. While a complete functional mapping of whole genomes is largely out of reach, a combination of theoretical modeling and new experimental tools may offer a new path to knowledge about global genomic organization. Here, two computer models were described that offer observables that are in principle testable through microarray expression studies.

One approach is based on the assumption that the evolution of gene regulatory circuits is governed by the requirement of robustness only. The resulting dynamics evolve networks which have a very large fraction of silent genes and short attractors. Thus they evolve to an ordered structure that counteracts the increasing chaos when networks become densely connected. The evolved architecture is characterized by simplicity of expression pattern and increased robustness to permanent mutational fluctuations in the network architecture – features that are also seen in real molecular networks. The second approach is based on local evolution of single gene function within a broad but otherwise unspecified window of average activity of this gene. This weak constraint of a selectively preferred activity range leads to a self-organized global activity of the artificial genetic network near a critical phase transition between the ordered and chaotic regime of this network.

Both models compare well with main features of current experimental data, but also pose new questions to experimentation that can only partly be answered with currently available data, but should easily be testable with current experimental techniques. In particular, a statistical approach to gene expression experiments, while feasible by new techniques, has only rarely been taken. Examples of experiments are knock-out experiments of random regulatory genes in order to test for average behavior as well as avalanches in changes of regulatory gene activity. Further, knock-out of random pairs of regulatory genes can serve to measure their correlation and provide important information about the global structure and organization of the genetic network. In particular, si-

multaneous knock-outs of random regulatory genes that belong to different co-expressed gene clusters will provide information about the hierarchical structure of the genome. The statistical distribution of all such experiments will provide new information about the global organization of gene regulation. With microarray techniques, the basis for such experiments is available.

References

Alon, U., Surette, M.G., Barkai, N., and Leibler, S. (1999). Robustness in bacterial chemotaxis. Nature *397*, 168 – 171.

Bak, P., and Sneppen, K. (1993). Punctuated equilibrium and criticality in a simple model of evolution. Phys. Rev. Lett. *71*, 4083 – 4086.

Bak, P., Tang, C., and Wiesenfeld, K. (1987). Self-organized criticality: an explanation of the 1/f noise. Phys. Rev. Lett. *59*, 381 – 384.

Bastola, U., and Parisi, G. (1996). Closing probabilities in the Kauffman model: an annealed computation. Physica A *98*, 1 – 25.

Bastola, U., and Parisi, G. (1998). Relevant elements magnetization, and dynamical properties in Kauffman networks: a numerical study. Physica D *115*, 203 – 218; The modular structure of Kauffman networks. ibid. 219 – 233.

Bhalla, U.S., and Iyengar, R. (1999). Emergent properties of networks of biological signaling pathways. Science *283*, 381 – 387.

Bittner, M., Meltzer, P., and Trent, J. (1999). Data analysis and integration: of steps and arrows. Nature Genet. *22*, 213 – 215.

Blattner, F. (2000). Blattner laboratory website. http:// www.genetics.wisc.edu/html/k12.html.

Bollobas, B. (1985). Random Graphs (London, UK: Academic Press).

Bornholdt, S., and Sneppen, K. (1998). Neutral mutations and punctuated equilibrium in evolving genetic networks. Phys. Rev. Lett. *81*, 236 – 239.

Bornholdt S., and Sneppen, K. (2000). Robustness as an evolutionary principle. Proc. R. Soc. London B *267*, 2281 – 2286.

Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., and Davis, R.W. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. Mol. Cell *2*, 65 – 73.

Christensen, K., Donangelo, R., Koiler, B., and Sneppen, K. (1998). Evolution of random networks. Phys. Rev. Lett. *81*, 2380 – 2383.

D`Haeseleer, P., Liang, S., and Somogyi, R. (2000). Genetic network inference: from Co-expression clustering to reverse engineering. Bioinformatics *16*, 707 – 726.

DeRisi, J.L., Iyer, V.R., Brown, P.O. (1997). Exploring the metabolic and genetic control on a genomic scale. Science *278*, 680 – 686.

Derrida, B. (1987). Dynamical phase transition in non-symmetrical spin glasses. J. Phys. A *20*, L721 – L725.

Derrida, B., and Pomeau, Y. (1986). Random networks of automata: a simple annealed approximation. Europhys. Lett. *1*, 45 – 49.

Dieckmann, U.M., and Doeberli, K. (1999). On the origin of species by sympatric speciation. Nature *400*, 354 – 357.

Elena, S.F., and Lenski, R.E. (1999). Test of synergistic interactions among deleterious mutations in bacteria. Nature *390*, 395 – 398.

Erdös, P., and Renyi, A. (1960). On the evolution of random graphs. Magyar Tud. Akad. Mat. Kut. Int. Kozl. *5*, 17–61.

Ewing, B., and Green, P. (2000). Analysis of expressed sequence tags indicates 35000 human genes. Nature Genet. *25*, 232–234.

Gould, S.J., and Eldredge, N. (1993). Punctuated equilibrium comes of age. Nature *366*, 223–227.

Hartwell, L. (1997). Theoretical Biology: a robust view of biochemical pathways. Nature *387*, 855–857.

Hertz, J. (1998). Statistical issues in reverse engineering of genetic networks. Pacific Symp. Biocomp. *98*, http://www.nordita.dk/hertz/papers/dgshort.ps.gz.

Holstege, F.C.P., Jennings, E. G., Wyrick, J. J., Lee, T. I., Hengartner, C. J., Green, M. R., Golub, T. R., Lander, E. S., and Young, R. A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717–728.

Kürten, K.E. (1988a). Correspondence between neural threshold networks and Kauffman Boolean cellular automata. J. Phys. A *21*, L615–L619.

Kürten, K.E. (1988b). Critical phenomena in model neural networks. Phys. Lett. A *129*, 157–160.

Kaneko, K., and Yomo, T. (1999). Isologous diversification for robust development of cell society. J. Theor. Biol. *199*, 243–256.

Kauffman, S.A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. *22*, 437–467.

Kauffman, S.A. (1990). Requirements for evolvability in complex systems: orderly dynamics and frozen components. Physica D *42*, 135–152.

Kauffman, S.A. (1993). The Origins of Order: Self-Organization and Selection in Evolution (Oxford, UK: Oxford University Press).

Kimura, M. (1983). The Neutral Theory of Molecular Evolution (Cambridge, UK: Cambridge University Press).

Lande, R. (1985). Expected time for random genetic drift of a population between stable phenotypic states. Proc. Natl. Acad. Sci. USA *82*, 7641–7644.

Laub, M.T., McAdams, H.H., Feldblyum, T., Fraser, C.M., and Shapiro, L. (2000). Global analysis of the genetic network controlling a bacterial cell cycle. Science *290*, 2144–2148.

Little, J., Shepley, D.P., and Wert, D.W. (1999). Robustness of a gene regulatory circuit. EMBO J. *18*, 4299–4307.

Newman, C.M., Cohen, J.E., and Kipnis, C. (1985). Neo-darwinian evolution implies punctuated equilibrium. Nature *315*, 400–402.

Newman, M.E.J., and Eble, G.J. (1999). Power spectra of extinction in the fossil record. Proc. R. Soc. London *B266*, 1267–1270.

Niehrs, C., and Pollet, N. (1999). Synexpression groups in eukaryotes. Nature *402*, 483–487.

Paczuski M., Bassler, K., and Corral A. (2000). Self-organized networks of competing Boolean agents. Phys. Rev. Lett. *84*, 3185–3188.

Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R.E., and Blot, M. (1999). Genomic evolution during a 10,000-generation experiment with bacteria. Proc. Natl. Acad. Sci. USA *96*, 3807–3810.

Patel, N.H., and Prince, V.E. (2000). Beyond the Hox complex. Genome Biol. *1*, 1027.1–1027.4.

Pennisi, E., and Roush, W. (1997). Developing a new view on evolution. Science *277*, 34–37.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Mannett, N., Kanin, E. *et al.* (2000). Genome-wide location and function of DNA Binding proteins. Science *290*, 2306–2309.

Savageau, M.A. (1971). Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. Nature *229,* 542–544.

Schuster, P. (1997). Landscapes and molecular evolution. Physica D *107*, 351–365.

Shapiro, J.A. (1998). Genome system architecture and natural genetic engineering in evolution. Ann. NY Acad. Sci. *870*, 23–25.

Sole, R.V. (1998). On Macroevolution, extinctions and critical phenomena. Complexity *1,* 40–44.

Sole, R.V., and Manrubia, S. (1997). Extinction and self-organized criticality in a model of large-scale evolution. Phys. Rev. E *54*, R42–R45.

Sole, R.V., Manrubia, S.C., Benton, M., and Bak, P. (1997). Self-similarity of extinction statistics in the fossil record. Nature *388*, 764–767.

Somogyi, R. and Sniegoski, C.A. (1996). Modelling the complexity of genetic networks: understanding multigenic and pleiotropic regulation. Complexity *1,* 45–63.

Stauffer, D., and Aharonyi, A. (1995). Introcution to Percolation Theory (Basingstoke Hauts, UK: Taylor Press).

Strogatz, S. (2001). Exploring complex networks. Nature *410*, 268–276.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. (1999). Systematic determination of genetic network architecture. Nature Genet. *22*, 281–285.

The Chipping Forecast (1999). Nature Genet. *21*, 1–60.

Thieffry, D., Huerta, A.M., Perez-Rueda, E., Collado-Vides, J. (1998). From specific gene regulation to global regulatory networks: a characterization of *Escherichia coli* transcriptional network. Bioessays *20* 433–440.

Van Valen, L. (1973). A new evolutionary law. Evolutionary Theory *1*, 1–30.

Wagner, A. (1996). Does evolutionary plasticity evolve? Evolution *50*, 1008–1023.

Wagner, A. (2000a). Decoupled evolution of coding region and mRNA expression patterns after gene duplication: Implications for the neutralist-selectionist debate. Proc. Natl. Acad. Sci. USA *97*, 6579–6584.

Wagner, A. (2000b). Robustness against mutations in genetic networks of yeast. Nature Genet. *24*, 355–361.

Wen, X., Fuhrman, S., Michaels, G.S., Carr, D.B., Smith, S., Barker, J.L., and Somogyi, R. (1998). Large-scale temporal gene expression mapping of central nervous system development, Proc. Natl. Acad. Sci. USA *95*, 334–339.

White, K.P., Rifkin, S.A., Hurban, P., and Hogness, D.S. (1999). Microarray analysis of *Drosophila* development during metamorphosis. Science *286*, 2179–2184.

Wright, S. (1982). Character change, speciation, and the higher taxa. Evolution *36*, 427–443.