

Evolution of Boolean Networks: Contingency Versus Necessity in Biology

Stephen Gilmore^{*1} and David Green²

¹*Dermatology Research Centre, School of Medicine, University of Queensland, Princess Alexandra Hospital, Brisbane, 4102, Australia*

²*Information Technology Services, University of Queensland, Brisbane, 4072, Australia*

Abstract: Many scientists are convinced that the Modern Synthesis – the current paradigm in evolutionary biology – needs to be expanded within the framework of an Extended Evolutionary Synthesis (EES). A major task for the EES is to provide an explanation for the origins and extant features of biological complexity. Here we address this issue, focusing our investigation on genetic network architectures, motifs, and dynamical behaviour, by developing an intuitive and essentially parameter-free evolutionary model of transcriptional regulation where the self-replicating digital organisms are Boolean networks, and where fitness is determined by their information-processing capacities. We validate our choice of fitness function by demonstrating that our evolved networks exhibit typical biological features of extant genetic regulatory networks: sparse connectivity, scale-free out-degree (within our range of measurement) and exponentially decaying in-degree distributions, significant clustering, a high proportion of feed-forward loop (FFL) network motifs, a prominence of canalising logic at the promoter, plasticity, and distributed robustness to mutation. In addition, the dynamics of our evolved networks feature simple attractor cycles that are robust to perturbations and exhibit self-organised criticality. In networks evolved without gene duplications, we show that the key architectural signatures noted above are absent. Surprisingly, the canalising fraction is much higher in comparison with networks evolved with gene duplications. These results suggest network properties of extant gene networks require gene duplications in order to evolve, and that these properties undergo positive selection, where they contribute to the global stability of the networks. By demonstrating that networks evolved without gene duplications are robust and, like their scale-free counterparts, also exhibit self-organised criticality, this work highlights the interplay between *contingent mechanism*, such as gene duplications, and *selection*, in determining evolutionary outcomes.

Keywords: Boolean networks, genetic networks, scale-free, robustness, evolvability.

INTRODUCTION

A major challenge for evolutionary theory is to explain how and why organisms develop complex features as a consequence of selection, mutation, and other non-selective forces [1,2,3]. Although the reconstruction of phylogenies by integrating information from palaeontology, ‘evo-devo’ and modern systems biology has met with partial success [4,5], efforts to trace population and molecular events in detail are undermined by a lack of historical data. Moreover, attempts to determine which features of biological systems perform vital roles, and which are incidental to physical or evolutionary mechanisms, are confounded by the singular origin of all extant life on Earth.

To be sure, the latter uncertainty is exemplified by considering the architectural signatures associated with genetic regulatory networks: why, for example, are out-degree (see Table 1 for a glossary of terms) distributions scale-free? Are these architectures performing an indispensable biologic function, or are they simply a relic secondary to the mecha-

nisms of genome expansion? While some authors have suggested such architectures confer biologic advantage, such as robustness or efficient information transfer [6,7,8], and therefore may be selected for, others have offered more pedestrian explanations: for example, it has been suggested that large-scale network structures can arise during neutral evolution [9], or can develop passively as a consequence of genome expansion *via* gene duplications [10,11,12]. Evolutionary modelling has also been invoked: previous work has shown that the development of scale-free architectures, perhaps not surprisingly, can develop in evolutionary runs associated with preferential attachment mutations [6,13,14]. At the intermediate level of gene network organisation, the preponderance of specific network motifs, such as the feed-forward loop motif [15], also demand explanation. While some authors have suggested they arise locally and in response to selective forces [16], others have argued, analogously to the argument associated with the development of scale-free degree distributions, that the frequency of such motifs can be explained by the mechanisms of genome expansion [17].

Previous work has shown many genes are regulated at the promoter by transcription factors in a combinatorial manner [16,18], and the logic has been shown to belong to a class of functions that possess the *canalising* property [18]. What is the origin and purpose of this class of combinatorial logic in

*Address correspondence to this author at the Dermatology Research Centre, University of Queensland, School of Medicine, Princess Alexandra Hospital, Brisbane, Queensland, 4102, Australia; Tel: + 61 7 3240 5467; Fax: + 61 7 3240 5663; E-mail: s.gilmore1@uq.edu.au

Table 1. Glossary of terms

| | |
|------------------------------|--|
| Cellular Automata | A grid of cells, connected to a variable number of nearest neighbours, where each cell can exist in two or more states, and where the state of each cell is updated as a function of the set of states of its connected nearest neighbours. |
| Random Boolean Network (RBN) | A generalised cellular automata where the architecture of connections between cells is random, and where the updating Boolean rules for each cell are chosen at random. |
| Connectivity | The density of connections between cells in a RBN, usually defined as the number of connections divided by the number of possible connections. |
| In-degree | The pattern of connectivity of an RBN – its architecture – is usually represented as a directed graph; that is, the output of one cell is input to another cell, not necessarily vice versa. The in-degree of a cell is its number of input connections. |
| Out-degree | The number of output connections from a given cell. |
| Clustering Coefficient | The likelihood that two connected cells are both connected to a third cell, averaged over all connected pairs. |
| Canalising Rules | A type of Boolean rule where the state of a cell is determined by the input from only one or more input cells, regardless of the states of the other input cells. |
| Attractors | Patterns of network states in an RBN that can be unchanging (fixed point), recurrent (oscillatory) or non-repeating (chaotic). |
| Avalanche Size | The number of iterations required for a RBN to reach an attractor following a single cell perturbation to an RBN settled on an existing attractor. |
| Self-organised Criticality | A system where the dynamics evolve toward a state where the distribution of avalanche sizes follows a power law. |
| Mutational Robustness | Fraction of non-viable offspring following random mutations; to reduce statistical error the robustness of a given network is calculated following 1000 independent random mutations. |
| Evolvability | Fraction of offspring with increased fitness following random mutations; to reduce statistical error the evolvability of a given network is calculated following 1000 independent random mutations. |
| Plasticity | Fraction of offspring with either no change or increased fitness following rewiring mutations. |

extant biological networks? Do they exist, as suggested by Kauffman [19], to sustain global ordering of the networks' behaviour, or, can their increased frequencies be attributed to additional or alternative explanations: for example, to facili-

tate mutational robustness? From the dynamical perspective, it has been shown by a number of investigators that the behavioural responses of genetic regulatory networks are critical: perturbations to gene expression lead to cascades of gene expression changes where the magnitudes of those changes follow a power law [20]. While it may seem likely that dynamic behaviour of this kind has evolved to facilitate robust yet flexible signalling responses, the origins of such cellular responses are less clear. An important question thus arises: are the scale-free architectures found in biological networks a necessary condition for critical dynamics to evolve?

Taken together, the examples above highlight an unresolved problem with regard to the structure and behaviour of biological networks: *contingency* versus *necessity*. In the following sections we develop an evolutionary model with the aim of gaining insight into this problem, paying attention to the development of architectural signatures, dynamical behaviour, robustness, and evolvability. We are motivated by the utility of evolutionary modelling: indeed, it is well recognised that a complementary approach to the investigation of life's evolutionary history is to study the evolution of digital organisms or *in-silico* genetic networks; the advantages of these methodologies have been described elsewhere [21,22].

Following Kauffman [19] and a large number of subsequent investigators, we model genetic regulatory networks with Boolean networks. A model genetic regulatory network based on Boolean logic is a type of generalised cellular automata, where the genes, represented by vertices in a directed graph, can be either in on or off states, corresponding to expression or repression of those genes. Gene states are updated simultaneously and discretely. Regulation at the promoter is modelled using Boolean logic: the state of vertex V_A at time $T = t + 1$ is determined by the state(s) of its input vertices (V_B, V_C, V_D, \dots) at time $T = t$, calculated using a fixed Boolean rule associated with vertex A . For example, a simple Boolean rule associated with gene A with inputs from genes B and C might be as follows: if any or both of the input genes B and C are expressed at $T = t$, then gene A is expressed at time $T = t + 1$, otherwise gene A is repressed at time $T = t + 1$. Fig. (1a) shows a five-vertex network, while Fig. (1b) shows a pattern of gene expression associated with this network. In Fig. (1b) time progresses down the page, and each gene is either expressed (yellow) or repressed (purple). Note that the state of the network at any time is represented by a horizontal row, and that this state repeats after $t = 4$ updates. The repeating pattern of network states forms an *attractor* of the network. In general, there can be many attractors associated with a given network (where the latter is defined by its architecture and list of associated Boolean rules). The *basin of attraction* of a given attractor is the set of states of the network that will evolve toward that attractor and stay there. Basins of attraction are not usually of equal size; from a random initial state, a network is more likely to evolve to particular attractors.

A Boolean network is then considered an 'agent'; to model the evolution of a population we track the changes in a cohort of networks over time. Briefly, each network among a population is assigned a fitness value based on its attractor pattern; at each replication round, individual networks are chosen randomly with a frequency *proportional to fitness*.

Table 2. Mutation-type look up table. Note how we have paid careful attention to the interplay between the types of mutations and their effects on the networks' architecture and logic. For example, while a mutation to a promotor will lead to a new Boolean rule at that promotor, a mutation to a transcription factor's gene will not lead to new logic at its target promotor; in such cases although the deletion of the corresponding edge requires a new bit length for the rule at that target promotor, it must be identical to the previous rule assuming absence of the lost transcription factor

| Mutation Type | Network Change | | Probability |
|--|---|--|-----------------|
| | Architectural Change | Rule Change | |
| <i>Promoter mutation</i> | No change | New random rule | 0.40 |
| | New edge | New random rule | 0.175 |
| | Deletion of edge | New random rule | 0.0875 |
| | Loss of all incoming edges | New random rule | 0.075 |
| <i>Transcription factor mutation</i> | Deletion of edge | Same rule as per pre-mutated transcription factor absent | 0.0875 |
| | Loss of all outgoing edges | Same rules as per pre-mutated transcription factor absent | 0.075 |
| <i>Duplication</i> | All outgoing edges as per duplicated vertex. New random incoming edge | Same rules for outgoing edges. New random rule for incoming edge | 0.02 |
| | All incoming edges as per duplicated vertex. New random outgoing edge | Same rules for incoming edges. New random rule for outgoing edge | 0.02 |
| <i>Horizontal t/fer</i> | | | |
| <i>a. External promoter mutation</i> | <i>a. New ingoing edge to new vertex</i> | <i>a. New random rule</i> | <i>a. 0.005</i> |
| <i>b. External transcription factor mutation</i> | <i>b. New outgoing edge from new vertex</i> | <i>b. Same rule as per new transcription factor absent. New random rule for new vertex</i> | <i>b. 0.005</i> |
| <i>Deletion</i> | Loss of all incoming and outgoing edges | Same rules as per deleted transcription factor absent | 0.05 |

interpret the sequence of specific states of the networks' functional attractors as coded environmental information. Since we do not simulate a single-goal optimisation process, we capture the essence of co-evolution and hence avoid the problem of coupling evolution and teleology. Finally, our fitness function has the realistic and desirable property of not corresponding to a well-defined metric in genome space.

Monte Carlo Updating Scheme

At each replication round, a network is chosen at random with a frequency *proportional to fitness* and replicates. The daughter network mutates with probability $p = mN / (1+mN)^4$ where m is the mutational burden and N is the vertex count. The types of mutations that can occur and their associated probabilities are shown in Table 2; although these rules and their frequencies are chosen heuristically, they are simple, biologically relevant [15], and, as noted above, unbiased.

If a mutation occurs, the daughter network's first attractor to be reached from random initial conditions is used to calculate its fitness, thus, as noted in the Introduction, we implicitly select for networks with highly non-uniform sized basins of attraction. To remain viable, the daughter network

is constrained to share greater than 90% attractor homology with its parent since gene expression patterns need to be similar otherwise they will not function as before. We apply our 90% attractor homology rule as follows: an attractor pattern can be written as a matrix of dimensions $N \times 256$, where N are the number of vertices and 256 the number of sequential network states sampled. When comparing attractors between the old network and the newly mutated network, we find an alignment between the two matrices that maximises the number of correct matches at all positions. The fraction of matching positions is then is a measure of attractor homology. Finally, the new network's attractor length is required to be less than 256 states. Fig. (1) shows an example of this process for a simple network of size $N = 5$. Irrespective of whether a mutation occurs, the daughter network then replaces another network within the population at random. If the mutated network fails to satisfy our constraints it is rendered non-viable and assumes a very low fitness. The constraints on viability ensure that the particular attractor patterns exhibited by our networks – which, as we noted above, are meaningful in the context of their environments – are preserved with high fidelity, and serve to exclude long period attractors, which could potentially include chaotic attractors, as biologically viable. Indeed, both long and shorter length chaotic attractors are not expected to evolve since the likelihood of a chaotic network mutating from either a cha-

⁴ This simple and monotonically increasing relation ensures $p < 1$ for any m and N .

otic or non-chaotic network with 90% or greater attractor homology is essentially zero. For the seed population where $N = 4$, calculating fitness by sampling over 256 iterations of its functional attractor ensures chaotic behaviour is associated with fitness values close or equal to zero.

Simulations

We begin all simulations with a seed population of 250 randomly generated 4-vertex networks and calculate the fitness for each. Although this procedure allows different networks and attractors to be viable, this variation is small compared with the space of all possible 4-vertex networks. For networks that grow from 4 to 50 vertices, our mutation probability expression p with $m = 0.0025$ yields a probability of mutation per replication event ranging from $p = 0.01$ to $p = 0.11$, consistent with a network module embedded within a primitive genome of 400 genes operating near its error threshold⁵ [32]. We therefore take $m = 0.0025$ as our baseline mutation load, and terminate our simulations when any member of the evolving population reaches $N = 50$. To begin, we run 28 independent and randomly seeded evolutionary histories from $N = 4$ to $N = 50$ at $m = 0.0025$. To investigate whether evolved network properties are augmented by further increases in network size, we take 7 of our $m = 0.0025$ networks at $N = 50$ and allow their evolution to continue, terminating the simulation when network growth became very slow. We then investigate the temporal evolution of network features by extracting the fittest network from 6 of our 28 $m = 0.0025$ evolutionary runs every 500 generations up to $N = 50$, and then, in some cases, continuing the evolutionary runs until generation 4500. To investigate how the properties of our evolved networks depend on the mutation load we run 8 independent randomly seeded evolutionary histories from $N = 4$ to $N = 50$ for $m = 0.04, 0.02, 0.01$ and 0.005 respectively. Finally, we use our model to perform three specific investigations: first, we run 9 randomly seeded evolutionary histories without gene duplications; second, we investigate network plasticity (the ability of a network to maintain or increase fitness following rewiring mutations) by measuring a single evolved network's average robustness following 1000 independent single random edge additions; and third, we investigate the fitness-robustness properties of two $N \sim 50$ populations derived from running our model for each of the different mutation rates noted above.

All features shown in Table 3 are calculated from the fittest network extracted from the population at the end of each of 28 evolutionary runs at $N = 50$. For the nine fitness measures, we take average values following 1000 repeated random mutations (of the particular type). For the dynamical stability investigations, we perform 1000 independent single-vertex random perturbations from random states on the networks' functional attractors, and determine the magnitudes of the respective subsequent avalanches.

RESULTS

We first investigate the characteristics of the fittest $N = 50$ networks of each of our 28 populations that evolved at m

Table 3. Network characteristics of the fittest network and their respective values calculated from 28 independent evolutionary simulations for $m = 0.0025$. See Fig. 2 for the distribution of values for features 2, 5, 6, 7, 8 and 9.

| Feature | | |
|---|--------|-------|
| | Mean | SD |
| 1. Size | 50.67 | 0.98 |
| 2. Fitness | 3824 | 720 |
| 3. Functional attractor length | 50.25 | 37.25 |
| 4. Connectivity | 0.05 | 0.005 |
| 5. In-degree | | |
| – maximum | 6.92 | 1.58 |
| – degree distribution p -value of log-log plot | 0.16 | 0.16 |
| – degree distribution regression coefficient of log-log plot | 0.63 | 0.18 |
| 6. Out-degree | | |
| – maximum | 12.4 | 3.21 |
| – degree distribution p -value of log-log plot | 0.01 | 0.01 |
| – degree distribution regression coefficient of log-log plot | 0.84 | 0.09 |
| – degree distribution scaling exponent | 1.48 | 0.13 |
| 7. Clustering coefficient | 0.15 | 0.04 |
| 8. Feed-forward loop motif count | 34.5 | 17.0 |
| 9. Fraction of canalising functions | 0.40 | 0.08 |
| 10. 1000 normal evolution frequency mutations | | |
| – fraction increased fitness | 0.23 | 0.10 |
| – fraction non-viable | 0.31 | 0.11 |
| – average fitness | 0.68 | 0.11 |
| 11. 1000 duplication / horizontal transfer mutations | | |
| – fraction increased fitness | 0.52 | 0.11 |
| – fraction non-viable | 0.20 | 0.09 |
| – average fitness | 0.76 | 0.09 |
| 12. 1000 deletion mutations | | |
| – fraction increased fitness | 0.02 | 0.02 |
| – fraction non-viable | 0.31 | 0.11 |
| – average fitness | 0.65 | 0.10 |
| 13. 1000 single vertex perturbations | | |
| – fraction new attractor | 0.20 | 0.07 |
| – average avalanche size | 4.60 | 1.99 |
| – avalanche magnitude distribution p value of log-log plot | 0.0002 | 0.001 |
| – avalanche magnitude distribution regression coefficient of log-log plot | 0.80 | 0.09 |
| – avalanche magnitude distribution scaling exponent | 1.60 | 0.42 |

⁵ For evolutionary systems, the approximate error threshold occurs when there is one mutation per organism per replication event. Higher mutation rates lead to irreversible information loss. For a primitive genome of 400 genes, the probability of mutation p at the error threshold for N embedded network modules' genes is given by $p = N / 400$. If $N = 4$, $p = 0.01$, and if $N = 50$, $p = 0.12$, as required.

Fig. (4). Evolved characteristics for a $N = 73$ network. Network graph (a) showing a highly connected region at top right. Histogram of in (b) and out (c) degree distributions reveal a fat tail for the out-degree distribution. This latter finding is reflected in (d), which shows a log-log plot of frequency versus outgoing connection counts. Over the scale of the plot, the relationship follows a straight line, with scaling exponent 1.33. This same relationship is evident in (e), where, following single vertex perturbation to random attractor states, the log-frequency of avalanche size is plotted against the log of avalanche size. Over the scale of measurement, the plot follows a straight line, indicating the dynamic behaviour of the network exhibits self-organised criticality. Here the scaling exponent is 1.71.

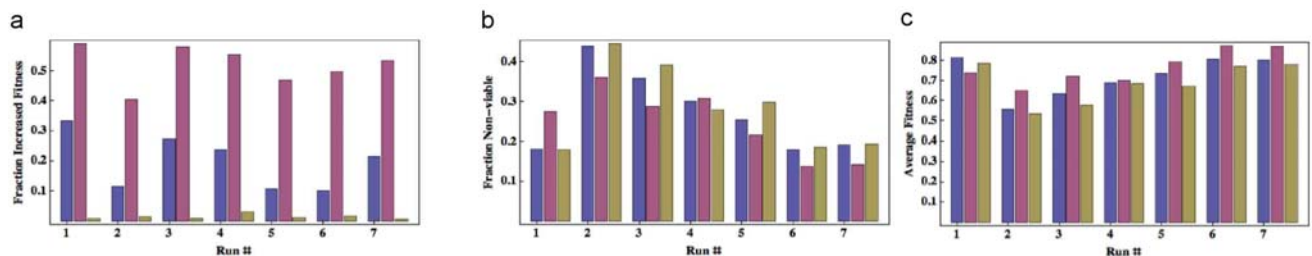


Fig. (5). Mutational robustness of networks for $N \sim 70$. Here we investigate the robustness of the 7 fittest networks extracted from 7 independently evolved populations at $m = 0.0025$. In (a), we represent evolvability by the fraction of 1000 random mutations that lead to increased fitness; in (b) we show the fraction of 1000 random mutations that lead to non-viability, while in (c) we show the average fitness of the network following 1000 random mutations. For each trio of columns, the first represents the frequency of mutation types as per the evolutionary run, while the second and third columns represent duplication / horizontal transfer and deletion mutations respectively.

lights the evolved robustness of these seven networks to mutations.

To investigate the temporal evolution of network features we sample the fittest member of 6 $m = 0.0025$ populations every 500 generations up to generation 4500; the important results are shown in Fig. (6). We note that while mutational robustness (the fraction of mutated networks that remain viable) increases slowly with time, significant robustness exists in small networks. Overall, evolvability (the fraction of mutated networks with increased fitness) remains unchanged as network size increases. Do any of the features found in Table 3 exhibit any dependence on the mutation load during evolution? We find that none of our network

features are influenced by varying mutation rates with the exception of fitness; in this latter case there is a mild decrease in fitness for $N = 50$ networks that evolve at mutation rates greater than baseline. Surprisingly, network robustness does not vary with mutation rate: evolvability, the non-viable fraction following mutations, and the average fitness following mutations does not change with mutation rate (data not shown).

We next perform three specific investigations: 9 evolutionary runs from $N = 4$ to $N = 50$ at $m = 0.0025$ without gene duplications (in this case the duplication probabilities shown in Table 2 are replaced by horizontal transfer), an investigation of network plasticity, and a detailed investiga-

tion of two $N = 50$ populations that we recover from our simulations at each mutation load reported above. Fig. (7) reveals that scale-free out-degree distributions are not present in networks evolved without gene duplications (a comparison between maximal out-degree; $p < 10^{-3}$, WRS). In addition, we find network clustering and the feed-forward loop motif count are much reduced in comparison to networks evolved with duplications ($p < 10^{-3}$, WRS); these results are similar to those expected from random graphs with the same connectivity. Fig. (7) also demonstrates an un-

expected finding: networks evolved without duplications – and hence without scale-free architectures – exhibit greater mutational robustness ($p < 10^{-2}$, WRS), and contain a very high fraction of canalising functions ($p < 10^{-2}$, WRS). We find that despite these differences, networks evolved without duplications – like their evolved-with-duplication counterparts – also evolve to a self-organised critical state.

We now demonstrate the plasticity of our evolved networks. Following 1000 random and independent edge additions to a single large network, we find that while 19% mu-

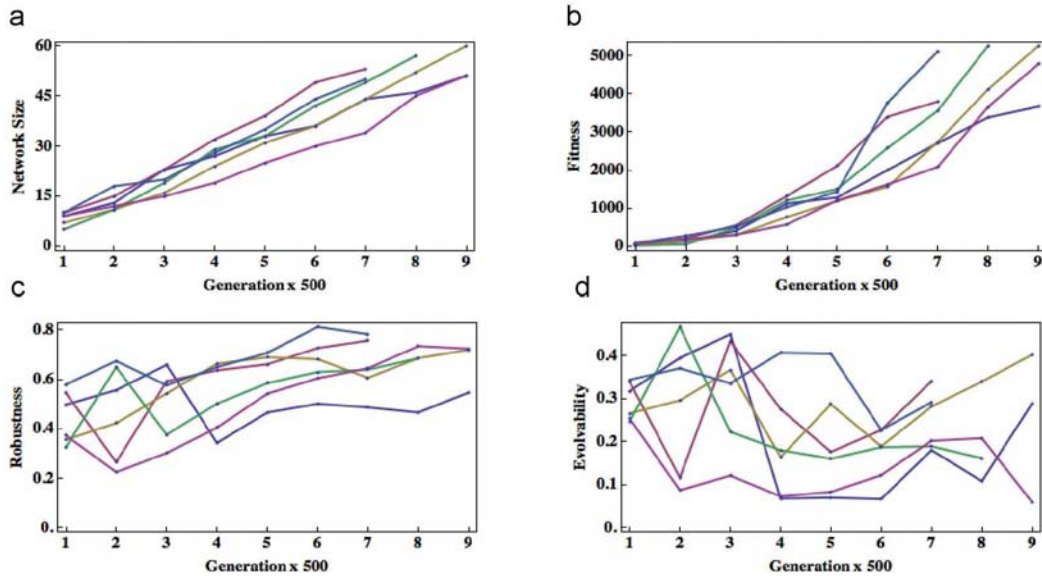


Fig. (6). Evolution of network features as a function of generation. Since our evolving populations contain 250 networks, each generation is equivalent to 250 Monte Carlo reproductive events, sampled from the population with a frequency proportional to fitness. Note fitness (a) grows faster than network size (b). Although robustness to mutation, on average, slowly increases with increasing complexity of the networks, one particular network was less robust compared with its much smaller ancestor. Evolvability does not increase during the evolution from very small to much larger networks.

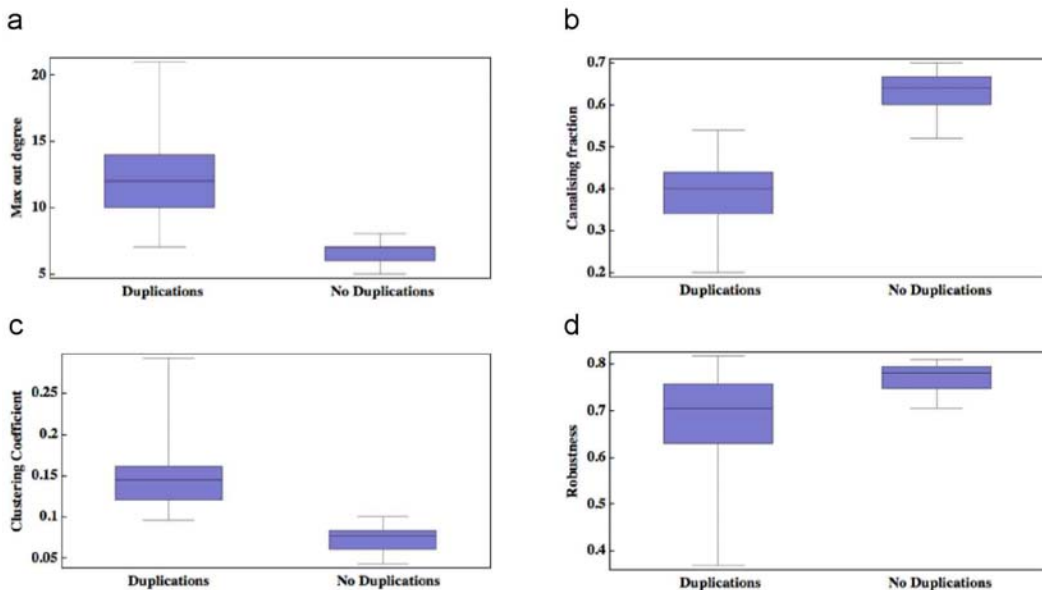


Fig. (7). Features of networks evolved with or without duplication events. Box-and-whisker plots demonstrating the differences between the two groups (duplications, $N = 28$, and no duplications, $N = 9$) with respect to maximal out-degree (a), canalising fraction (b), clustering coefficient (c) and mutational robustness (average fitness following 1000 random mutations, expressed as a fraction) (d). The differences between the two groups are all significant: $p < 0.001$ (a-c); $p < 0.003$ (d) (Wilcoxon rank sum test).

tants are non-viable, the remainder are associated with an average fitness equal to 98% of the network's initial fitness. Moreover, 6% of mutants exhibit a gain in fitness. These results, which are strikingly similar to those obtained with *E. coli* [33], indicate that this network is highly evolvable.

Finally, we investigate the correlations between fitness and mutational or dynamic robustness of all members belonging to two $N = 50$ populations evolved at each of the five mutation rates noted above. For nine populations, we find no evidence of intra-population correlations. One population demonstrates a correlation between fitness and mutational fragility; this result is shown in Fig. (8a) (linear regression, $p < 10^{-2}$). There is no evidence, however, that the distribution of members in any of our ten populations is evolving toward the *efficient frontier*, a hypothetical boundary in fitness-robustness phase space [34].

DISCUSSION

Here we establish the validity of our model by demonstrating that the important large and small-scale architectural features of extant gene regulatory networks can be reproduced in an evolutionary setting that lacks any predetermined bias for the development of these features [16,21,35]. Our networks are evolvable, are robust to mutation, and their functional attractors are generally short (see Table 3), a result consistent with the observed simplicity of single-gene switching events in eukaryotes [36]. By reproducing the *E. coli* re-wiring experiment of Isalan *et al.*, [33], we find similar quantitative results: our networks, like the genetic regulatory network in *E. coli*, are *plastic*; that is, they can tolerate network re-wiring changes well, and will frequently benefit from such changes.

Our networks also exhibit self-organised criticality, a feature that evolves gradually with network growth, and is considered a general feature of cellular dynamics [20]. In the Boolean representation of genetic regulatory networks, although 'edge of chaos' dynamics are generally thought to require large frozen components [19,21], our networks are associated with attractor patterns where the majority of vertices oscillate. This contrary finding has important implications for the evolvability of networks [37]: homeostatic yet flexible dynamics can exist *without functional redundancy*. While it has been suggested that scale-free architectures may facilitate a reduction in functional redundancy [38], recent work suggests that criticality may maximise both robustness and evolvability [39]. Our results support all three of these contentions: first, since our fitness function implicitly selects for networks with low functional redundancy, we might expect all our networks to exhibit criticality: indeed, this is the case; and second and third, since our networks are critical even when small, we cannot refute the idea that this dynamic signature evolves readily in order to enhance (or, possibly, maximise) both robustness and evolvability.

Real-World Fitness

Does the success of our model in reproducing biologic network structure imply that within-network information processing is an important fitness measure in real biological systems? Although we have drawn from a large body of

work to justify our fitness measure (see Methods), clearly the abstract nature of the mutual information content of gene expression patterns implies that it cannot, on its own, be the sole determinant of fitness. Details matter, and although we incorporate semantic details in our model by constraining the fidelity of reproduction, mutations in real biological systems may at times be associated with increased fitness and a *reduction* in functional complexity. One possible solution to this shortcoming would be to introduce an evolutionary algorithm where certain attractor patterns near to the current attractor⁶ could be assigned a higher fitness, independent of the magnitude of the mutual information associated with the new attractor. However, giving greater fitness weightings to specific patterns (analogously to pattern-matching evolutionary algorithms) introduces teleology: evolution is constrained and becomes predictable, a less than desirable outcome. In any case, our overarching goal is not to gain insight with regard to real-world fitness measures; rather, we are motivated to develop a model that allows us to gain insight into the evolution of network architectures and dynamics. It is this latter consideration to which we now turn.

The Origin of Large and Small-scale Network Structure

While the origins and functions of extant biological gene regulatory network characteristics remain uncertain, there have been many investigations, both experimental and theoretical, that have attempted to address this issue, yielding a variety of explanations. In particular, the origins of scale-free out-degree distributions in biologic networks have been attributed, by various authors, to maximising information flow [7,8], to facilitate mutational robustness [6,40], as a consequence of neutral selection [9], and finally, as a by-product of gene duplications during evolution [10,11,12]. The latter explanation is still controversial; evidence obtained from the investigation of paralogous genes have also been used to argue that gene duplications are not responsible for the development of scale-free networks [41]. It is only the first explanation – maximising information flow within networks – that dovetails with the intuitive idea that selection effects shape biological phenotypes.

By demonstrating that networks grown without duplication mutations do not exhibit any of the typical architectural signatures found in extant genetic regulatory networks, we provide evidence that suggests duplications are an essential prerequisite for the evolutionary development of such features. Furthermore, the negative correlation between the maximal out-degree and the fraction of canalising inputs (demonstrable in our 28 networks grown from $N = 4$ to $N = 50$ (Fig. 8b) as well as in the comparison between networks grown with or without duplications (Fig. 7a, b) suggests – if we assume canalising inputs act to stabilise network dynamics at criticality – that scale-free architectures also contribute to the overall stability of the networks' dynamics. This result suggests that in biological systems, scale-free out-degree distributions are subject to *positive* selection. Indeed, the notion that positive selection for scale-free out-degree distributions may occur is augmented by a second line of evidence: when a hub regulates many genes, we may expect those regulated genes to regulate a number of common genes

⁶ One method of quantifying the 'nearness' of two attractor patterns in a cellular automata is by calculating the Hamming distance.

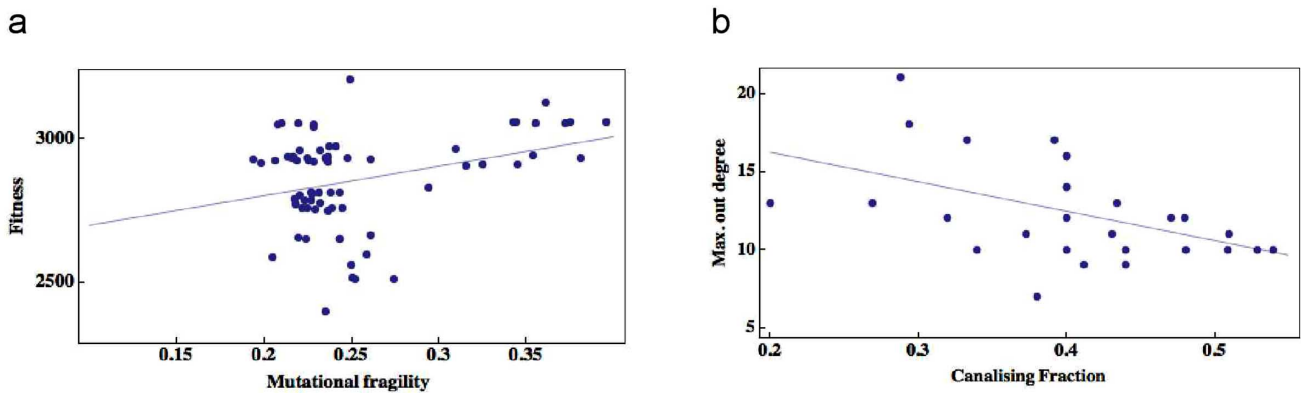


Fig. (8). Intra and inter-population correlations. Fitness-robustness phase space (a) for a population of $N = 50$ networks evolved at $m = 0.0025$. Each point represents identical members of the population. Note the correlation between mutational fragility and fitness, and the absence of evidence for the aggregation of points around the efficient frontier, a hypothetical boundary that would trace a line upward and to the right. For our 28 populations evolved at $m = 0.0025$, there exists a negative correlation between the maximal out-degree and the canalising fraction (b). Results shown in Fig. 7a, b are strongly suggestive of a similar interpretation.

since it is likely the duplicated genes are those that the hubs regulate. However, when we examine our networks with respect to the genes regulated by the genes that are in turn regulated by the hubs, we find no evidence for gene commonality; indeed, all the second-tier regulated genes are equivalent to a cohort of genes selected at random. This result indicates that our evolved networks are subject to significant rewiring following duplication events, and implies the maintenance of hubs is an architectural feature subject to positive selection. While the authors of two papers advocating duplications as a cause of scale-free network architectures argue against positive selection, they do not rule out this possibility [11,12].

Perhaps surprisingly, our networks evolved without duplications – and therefore without scale-free out-degree distributions, prominent clustering or increased numbers of feed-forward loop motifs – perform just as well (or better) than networks evolved with duplications: in particular, mutational robustness is enhanced (Fig. 7d). The networks also exhibit self-organised criticality, indistinguishable from their scale-free counterparts. These results suggest evolution could have led to very different network structures, demonstrate that scale-free architectures are not a necessary prerequisite for the evolution of critical dynamics, and imply that gene duplications are an example of *contingent mechanism*. Our results allow us to draw some tentative biological conclusions: the molecular machinery that allows segments of DNA to be duplicated is not then a prerequisite for fully functioning networks to evolve, but since it occurs, it leads to certain architectural features – scale-free out-degree distributions, network clustering and increased numbers of feed-forward loop motifs – of which the former, and probably the latter two, are then subject to positive selection.

The Evolution of Network Robustness and Evolvability

In contrast to the paradigm that scale-free architectures confer mutational robustness to networks, we have found that scale-free architectures are less mutationally robust than their counterparts evolved without gene duplications. We expect this result is due to the higher canalising fraction in the networks evolved without duplications, potentially augmented by the susceptibility of networks with hub genes to

be subject to catastrophic failure. We conclude that there is a significant trade-off between the maximal out-degree and the fraction of canalising inputs with respect to the mutational robustness of evolved networks, a result that suggests canalising inputs do not just stabilise global dynamics, but play a significant role in conferring mutational robustness. The interplay between network architecture and logic implies that the properties of network architectures, if examined in isolation, may yield incorrect conclusions with regard to their functional significance.

Is robustness a key driver for the evolution of complex features? While a previously reported conclusion based on the results of a fitness-free evolutionary model suggests that selection for robustness may play a non-trivial role in shaping network architecture and dynamics [21], other authors have suggested selection for robustness may facilitate network growth [42,43]. Selection for robustness is likely to play an important role in the evolution of the phage lambda genetic regulatory circuit [44], and may be analogous to ‘survival of the flattest’ [45]. Many of the above findings are perhaps not surprising. After all, robustness is, in a sense, the ultimate basis for fitness. We might think of robustness as the stage on which the ballet of neutral and fitness-driven evolution is played out. Many fundamental properties of gene network architecture and logic are likely to be determined by an absolute requirement for mutational robustness. But the evolutionary relationships between network features that confer robustness on one hand, and fitness on the other, are less clear. If robustness were a key driver of the evolution of the complex features that in turn augment network fitness then we would expect that, within evolving populations, a positive correlation between fitness and robustness should exist, and very few, if any, members of populations with both low fitness and low robustness should be present. On the other hand, if robustness drives the evolution of complex features that are deleterious to fitness, we would expect a within-population trade-off between fitness and robustness. This latter possibility is analogous to the idea of an efficient frontier, noted previously and introduced by Kitano, where such a trade-off defines a boundary in fitness-robustness phase space that members of populations cannot cross [34].

Yet there exists a third possibility: complex features conferring robustness to networks may only have marginal effects on fitness – they are essentially neutral. What do our results suggest? First, with the exception of one population evolved at $m = 0.0025$ (Fig. 8a), none of our other nine intra-population fitness-robustness scatter plots provide any evidence for fitness-robustness correlations, or evidence that the populations are evolving toward the efficient frontier. Second, we cannot demonstrate any inter-population correlations between fitness and robustness when we examine our 28 independent simulations at $m = 0.0025$. Finally, we have shown that significant robustness is already present in small networks (Fig. 6c), and that evolvability does not increase with increasing network complexity (Fig. 6d). Taken together, this line of evidence suggests robustness and evolvability are not major drivers for the evolution of the complex features associated with either fitness gains or losses. Our results thus support the neutral hypothesis, Fig. 8a notwithstanding. It is important to note that the term ‘complex feature’ is not precise; in this regard we suggest the existence of robustness and evolvability in networks may be due to *subtle* large or small-scale architectural organisation – complex features, nonetheless – that only have marginal effects on fitness and hitherto remain unidentified.

CONFLICT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

None declared.

REFERENCES

- [1] Pigliucci, M. Do we need an extended evolutionary synthesis? *Evolution*, **2007**, *61*: 2743-2749.
- [2] Carroll, R.L. Towards a new evolutionary synthesis. *Trends. Ecol. Evol.*, **2000**, *15*: 27-32.
- [3] Avise, J.C.; Ayala, F.J. In the light of evolution I: Adaptation and complex design. *Proc. Natl. Acad. Sci. USA*, **2007**, *104* (Suppl 1), 8563-8566.
- [4] Koentges, G. Evolution of anatomy and gene control. *Nature*, **2008**, *451*, 658-663.
- [5] Poelwijk, F.J.; Kiviet, D.J.; Weinreich, D.M.; Tans, S.J. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature*, **2007**, *e 445*, 383-386.
- [6] Barabasi, A.L.; Albert, R. Emergence of scaling in random networks. *Science*, **1999**, *286*, 509-512.
- [7] Braha, D.; Bar-Yam, Y. Topology of large-scale engineering problem-solving networks. *Phys. Rev. Stat. Nonlin. Soft-Matter.Phys.* **2004**, *69*, 016113, 1-7.
- [8] Sole, R.V.; Valverde, S. Information theory of complex networks: On evolution and architectural constraints. *Santa Fe Institute Working Paper*, 03-11-061.
- [9] Lynch, M. The evolution of genetic networks by non-adaptive processes. *Nat. Rev. Genet.*, **2007**, *8*: 803-813.
- [10] Gelfand, M.S. Evolution of transcriptional regulatory networks in microbial genomes. *Curr. Opin. Struct. Biol.*, **2006**, *16*: 420-429.
- [11] Amoutzias, G.R.; Robertson, D.L.; Oliver, S.G.; Bornberg-Bauer, E. Convergent evolution of gene networks by single gene duplications in higher eukaryotes. *EMBO Rep.*, **2004**, *5* (3), 274-279.
- [12] van Noort, V.; Snel, B.; Huynen, M. The yeast co-expression network has a small-world scale free architecture and can be explained by a simple model. *EMBO Rep.*, **2004**, *5* (3), 280-284.
- [13] Vazquez, A.; Rammini, A.; Maritan, A.; Vespignany, A. Modeling of protein interaction networks. *Complexus*, **2003**, *1*: 38-44.
- [14] Foster, D.V.; Kauffman, S.A. Network growth models and genetic regulatory networks. *Phys. Rev. Stat. Nonlin. Soft-Matter.Phys.*, **2006**, *73*: 031912, 1-8.
- [15] Babu, M.M.; Teichmann, S.A.; Aravind, L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J. Mol. Biol.*, **2006**, *358*: 614-633.
- [16] Babu, M.M. Structure, evolution and dynamics of transcriptional regulatory networks. *Biochem. Soc. Trans.*, **2010**, *38*: 1155-1178.
- [17] Cordero, O.; Hogeweg, P. Feed-forward loop circuits as a side-effect of genome evolution. *Mol. Biol. Evol.*, **2006**, *23*(10): 1931-1936.
- [18] Harris, S.E.; Sawhill, B.K.; Wuensche, A.; Kauffman, S. A model of transcriptional regulatory networks based on biases in the observed regulation rules. *Complexity*, **2002**, *7* (4): 23-32.
- [19] Kauffman, S.A. *The Origins of Order. Self-organization and selection in evolution*. Oxford University Press: New York, USA, Oxford, UK, **2003**.
- [20] Ramo, P.; Kesseli, J.; Yli-Harja, O. Perturbation avalanches and criticality in gene regulatory networks. *J. Theor. Biol.*, **2006**, *242*: 164-170.
- [21] Bornholdt, S. Modelling genetic networks and their evolution: A complex dynamical systems perspective. *Biol. Chem.*, **2001**, *382*: 1289-1299.
- [22] Lenski, R.E.; Ofria, C.; Pennock, R.T.; Adami, C. The evolutionary origin of complex features. *Nature*, **2003**, *423*: 139-144.
- [23] Carroll, S.B. *Endless forms most beautiful*. Phoenix: London, UK, **2005**.
- [24] Ay, N.; Olbrich, E.; Bertschinger, N.; Jost, J. A unifying framework for complexity measures of finite systems. *Santa Fe Institute Working Paper*, 06-08-028.
- [25] Komili, S.; Silver, P.A. Coupling and co-ordination in gene expression processes: a systems biology view. *Nat. Rev. Genet.*, **2008**, *9*: 38-48.
- [26] Arendt, D. A kernel bears fruit. *Science*, **2006**, *314*: 1085-1086.
- [27] Lozada-Chavez, I.; Janga, S.C.; Collado-Vides, J. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic. Acids. Res.*, **2006**, *34*: 3434-3445.
- [28] Wagner, A. Does evolutionary plasticity evolve? *Evolution*, **1996**, *50*: 1008-1023.
- [29] Quayle, A.P.; Bullock, S. Modelling the evolution of genetic regulatory networks. *J. Theor. Biol.*, **2006**, *238*: 737-753.
- [30] Crombach, A.; Hogeweg, P. Evolution of evolvability in gene regulatory networks. *PLoS Comp. Biol.*, **2008**, *4*: e1000112.
- [31] Adami, C. *Introduction to artificial life*. Springer-Verlag: Berlin, **1998**.
- [32] Mushegian, A.R.; Koonin, E.V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. USA*, **1996**, *93*: 10268-10273.
- [33] Isalan, M., et al. Evolvability and hierarchy in rewired bacterial gene networks. *Nature*, **2008**, *452*: 840-845.
- [34] Kitano, H. Violations of robustness trade-offs. *Mol. Syst. Biol.*, **2010**, *6*: 384.
- [35] Alon, U. *An introduction to systems biology: Design principles of biological circuits*. Chapman Hall / CRC: Boca Raton, Florida, USA, **2007**.
- [36] Cho, R.J.; Campbell, M.J.; Winzler, E.A.; Steinmetz, L.; Conway, A.; Wodicka, L.; Wolfsberg, T. G.; Gabrielian, A. E.; Landsman, D.; Lockhart, D. J.; Davis, R.W. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell*, **1998**, *2*: 65-73.
- [37] Pigliucci, M. Is evolvability evolvable? *Nat. Rev. Genet.*, **2008**, *9*: 75-82.
- [38] Fortuna, M.A.; Melian, C.J. Do scale-free regulatory networks allow more expression than random ones? *J. Theor. Biol.*, **2007**, *247*: 331-336.
- [39] Aldana, M.; Balleza, E.; Kauffman, S.; Resendiz, O. Robustness and evolvability in genetic regulatory networks. *J. Theor. Biol.*, **2007**, *245*: 433-448.
- [40] Wagner, A. Robustness against mutations in genetic networks of yeast. *Nat. Genet.*, **2000**, *24*: 355-361.
- [41] Teichmann, S.A.; Babu, M.M. Genetic network growth by duplication. *Nat. Genet.*, **2004**, *5*: 492-496.
- [42] Csete, M.E.; Doyle, J.C. Reverse engineering of biological complexity. *Science*, **2002**, *295*: 1664-1669.
- [43] Soyer, O.S.; Bonhoeffer, S. Evolution of complexity in signalling pathways. *Proc. Natl. Acad. Sci. USA*, **2006**, *103* (44), 16337-16342.

[44] Little, J.W.; Shepley, D.P.; Wert, D.P. Robustness of a gene regulatory circuit. *EMBO J*, **1999**, *18* (15): 4299-4307.

[45] Wilke, C.O.; Wang, J.L.; Ofria, C.; Lenski, R.E.; Adami, C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, **2001**, *412*: 331-333.

Received: November 17, 2011

Revised: February 28, 2012

Accepted: February 29, 2012

© Gilmore and Green.; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.