The Evolution of Evolvability in Gene Transcription Networks

Eduardo J. Izquierdo[1,2] and Chrisantha T. Fernando[3,2]

[1]Centre for Systems Biology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[2]Centre for Computational Neuroscience and Robotics, University of Sussex, Brighton BN1 9QH, UK
[3]Mathematical Biology, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK
eji21@sussex.ac.uk, ctf20@sussex.ac.uk

Abstract

We present a case of a genotype-phenotype map, that when evolved in variable environments optimizes its genetic representation to structure phenotypic variability properties, allowing rapid adaptation to novel environments. How genetic representations evolved is a relatively neglected topic in evolutionary theory. Furthermore, the "black art" of genetic algorithms depends on the practitioner to choose a representation that captures problem structure. Nature has achieved remarkably efficient heuristic search mechanisms without top-down design. We propose that an important example of this, ubiquitous in biology is the structuring of the phenotypic variability properties of gene networks. By studying a simple model of gene networks in which topology is a function of interactions between transcription factor proteins and transcription factor binding sites (TFBS), we show that transcription factor binding matrices (TFBM) evolve to positively constrain phenotypic variability in response to transcription factor binding sequence mutations.

Introduction

Where there is redundancy in the genotype to phenotype map, there is neutrality. For a given phenotype, if the distribution of phenotypes accessible by one-mutant neighbours differs depending on the particular genotype that encodes the initial phenotype, then there is non-trivial neutrality (Toussaint, 2003), meaning that there is variation in the phenotypic exploration distribution. Toussaint has shown that selection can act on the effective fitness of exploration distributions (i.e. the quasispecies fitness), and claims that this is the mechanism for the evolution of evolvability (Wagner and Altenberg, 1996). Evolvability is the capacity to rapidly adapt to novel environments by natural selection.

Add to this that the environment of the offspring differs from the environment of the parent, due to mutation and external variations, e.g. bacteria may sometimes be in the gut, and at other times outside the host (Ciliberti et al., 2007). Co-evolution is the norm, rather than the exception, meaning that the phenotype required for 'optimum' fitness is always changing, i.e. the fitness landscape fluctuates. We show that a fluctuating fitness landscape selects for exploration distributions with greater evolvability.

What mechanisms are capable of "biasing the kind and amount of phenotypic variation produced in response to random mutation, such that more favourable and non-lethal kinds of variation are available on which natural selection can act" (Kirchner and Gerhart, 1998). We demonstrate that a ubiquitous developmental mechanism, the formation of gene networks based on TFBM-TFBS interactions, has the capacity to allow heredity of exploration distribution variants in gene network topology space, and that this is likely to be the case in real gene networks.

A notable limitation of evolutionary algorithms is that the variational machinery is not self-referentially encoded[1], whereas in embodied evolution (because of the *necessity* of a developmental decoding of the genotype to produce the phenotype) the genetic representation (Toussaint, 2003), or genotype-phenotype map (Wagner and Altenberg, 1996), can maintain variants in the exploration distribution that can be acted upon by positive selection. Reisinger and Miikkulainen (2007) claim to have developed an indirect encoding for a neural network capable of structuring the genotype-phenotype map, to speed up the evolution of Nothello players. Other compact and indirect encodings of neural networks are effective in some specific problem domains (Gruau et al., 1996) , (Hornby and Pollack, 2002). Understanding the principles of network evolvability in natural systems is an important goal.

Defining Evolvability and Robustness

Several definitions of evolvability and robustness exist in the literature. We define evolvability as an ordering of the rate of evolution, between individuals of equal fitness, when exposed to directed selection from a starting point, S, to an end point, F, in phenotype space, where S is not equal to F. One could say that an individual A is more evolvable than individual B (from S to F) if the best offspring of A is on average fitter than the best offspring of B (Turney, 1987). Another quantitative measure of the evolvability of a variational system is the probability that an offspring has fitness

[1]Self-referential encoding refers to a genotype-phenotype map that is capable of non-trivial neutrality.

greater than the parent (Barnett, 2003). Given a probability density function of offspring fitnesses from a single parent, the evolvability of that parent is the fraction of offspring fitter than itself (Smith et al., 2002; Altenberg, 1994a)[2]. Note that with these definitions we can only unbiasedly compare evolvability between individuals of the same fitness. Also, there is no such thing as evolvability when $S = F$, since there is no directed selection, and so one cannot measure a rate of evolution. On the other hand, in the case where $S = F$, robustness *is* defined. It is a measure of the capacity of phenotypes to remain unchanged, given stabilizing selection. Evolvability and robustness are measures of evolutionary behaviour. They are emergent properties of exploration distributions (McGregor and Fernando, 2005).

Evolvability Sustaining Mechanisms

Kirchner and Gerhart (1998) have described various mechanisms that increase the probability that an offspring, varying within certain bounds, will be viable. One important class of these is exploration and exploitation mechanisms. For example, the immune system adapts to evolutionarily novel antigens by implementing somatic selection. Microtubules control and manipulate cell organelles and chromosomes, independently of the number and location of these items, thus allowing variation in these items to be viable, with respect to mitosis for example. Pathfinding by axonal growth cones allows neural structures to evolve, and still be viable (Kirchner and Gerhart, 1998). These search mechanisms allow robustness as well as evolvability (Wagner and Altenberg, 1996).

Another mechanism that confers evolvability and robustness is weak interaction. Kirchner and Gerhart (1998) contrast the complex transcription regulation of eukaryotes with the simple regulation of prokaryotes. Eukaryotes have complex cis-regulatory regions whereas prokaryotes do not. Eukaryotes have enhancer-binding proteins with limited affinity and low sequence specificity for enhancer sites. Their binding affinities may also be contingent on other proteins (non-independent site affinities). This property is called *weak-linkage*. Several authors have considered the role of weak interactions. It has been hypothesised that *weak interactions* confer evolvability (Conrad, 1990), and robustness (Volkert and Conrad, 1998). For example, Kirchner and Gerhart (1998) claim that Calmodulin is a *versatile inhibitor*, meaning that with few mutational steps it can bind to a protein for which it is selected to bind, attributing this to its "low sequence requirements" for binding to targets, that result from its flexibility and stickiness.

The above mechanisms *extend* the viable range of the exploration distribution, rather than constraining its directionality as emphasized in (Arthur, 2004). The exploration distribution can also make sense of another class of mechanism prevalent in bacteria. These are the mechanisms that maintain genotypic diversity in the population, increasing the chance that at least some of the existing variants will be pre-adapted to the new environment. This is a kind of bet-hedging. Typically, *E. coli* isolated from populations in the wild contain a small proportion with a 100 fold increased mutation rate due to inactivation of an error correcting enzyme (Matic et al., 1997). This proportion is much greater than expected in the absence of selection (Tenaillon et al., 2001). The full gamut of genetic and epigenetic devices for structuring variation (in terms of rate, site and inducibility) is discussed in (Rando and Verstrepen, 2007). Remarkably, Kussell and Leibler (2005) have shown that where the cost of maintaining diversity is less than the cost of sensing the environment, stochastic switching is selected over more complex sensing and response mechanisms. This is often the case in bacteria. The exploration distribution is skewed by such processes.

There are clear examples of highly conserved core processes that have optimized exploration distributions. According to a model by Zhu and Freeland (2006) the genetic code is optimized to allow the rapid adaptive evolution of proteins. Using a simple model of a sequence-to-protein-structure map, they mutated the sequence thus altering the structure. They used a genetic algorithm to re-evolve the original structure and found that with the existing genetic code, the structure could re-evolve much faster than if a randomly chosen code was used. Protein stability has also been argued to be an adaptation for evolvability. Bloom et al. (2006) showed using a lattice protein model that "extra stability is neutral with respect to selection for protein function, but it can be crucial in allowing a protein to tolerate [destabilizing] mutations that confer beneficial phenotypes". That is, a protein with more stability was able to evolve to a new desired function faster.

Modularity of various forms appears to underlie many kinds of evolvability (Wagner and Altenberg, 1996; Force et al., 2005; Lipson et al., 2002), since adaptation to environment A can be carried out without interfering with adaptation to environment B. At the cellular level, there may exist exotic exploration distribution structuring mechanisms that remain mysterious. For example, the pattern of gene expression in the lifetime of a paramecium influences which genes are passed to its offspring in a very complicated way (Prescott and Rozenberg, 2002).

Cognition is the cherry on the cake of exploration distribution structuring systems. The effectiveness of lifetime variation generation mechanisms tends to increase over evolutionary time, with solutions being transmitted across generations in novel ways. Many aspects of cultural inheritance, permitted by human thought and language, are adaptations that allow rapid adaptation to novel environments. Cognitive

[2]There are many other definitions of evolvability with different emphasis, e.g. "evolvability is the ability of the genetic system to produce *and maintain* potentially adaptive genetic variants" (Hansen, 2006).

mechanisms for generalization such as associative learning and symbolic reasoning allow us to choose behaviours with remarkable directedness compared to random search.

The Evolution of Evolvability Sustaining Mechanisms

How did such mechanisms evolve? Toussaint (2003) describes that selection pressure could act on *effective fitness*. To produce robust encodings Toussaint selects explicitly for neutral variants, a process that is intended to mimic stabilizing selection. This corresponds to Kirchner and Gerhart (1998) who say that evolvability is a by-product of selection for robust development in the face of internal (mutational) and external (environmental) noise. Secondly, adaptations for evolvability may have been selected because they allowed better exploration of new environments by clades. Hitchhiking of evolvability conferring genes along with advantageous traits whose appearance they facilitate is one mechanism that could achieve this (Conrad, 1990). Recently, Earl and Deem (2004) have shown in a model of protein evolution that the rate of mutation and the "swapping" of protein modules increases in variable environments to confer greater evolvability. Another proposed mechanism is constructional selection where selection acts to filter new loci. Alleles at new loci that have low epistasis are favoured because they are less likely to be fatal, resulting in modular genotype-phenotype mappings Altenberg (1994b). Finally, Kashtan and Alon (2005) have shown that selection in variable environments with modular goals results in the establishment of an intermediate genotype state that can rapidly mutate to become optimal in either environment.

We demonstrate that adaptations for evolvability in gene networks arise due to individual level selection in variable environments. By using a genetic algorithm to evolve agents under fixed versus variable environments, we identify how exploration distributions are restructured by TFBM evolution, a process that is also expected in natural evolution.

The Construction of Gene Networks

Gene network topology emerges from the interaction of transcription factors (TFs) binding to "degenerate families" of transcription factor binding sites (TFBSs) that are of 5-25 nucleotides in length, situated on promotors (Moses et al., 2003). Degenerate refers to the fact that different transcription factor binding sites that bind the same TF protein may differ in 20-30 percent of bases (Collado-Vides et al., 1991). In *E. coli*, promotors are approximately 500 base pairs long and contain several TFBSs (Berg et al., 2004). A position-weight matrix (or transcription factor binding matrix, TFBM) represents the binding preferences of a TF. Empirically these can be inferred from genome sequences if independent evidence of TF binding exists (Stormo, 2000). The binding energy between a TF and the TFBS can be well approximated by the sum of independent contributions from positions in the binding site.

What are the modes of evolution of gene regulatory networks? Gelfend (2006) discusses the various approaches to this question. Many phenotypic differences between species are attributable to changes in gene expression patterns rather than changes in structural or metabolic proteins (Tirosh et al., 2008). How does evolution of the gene regulatory network take place? Babu et al. (2006) have shown that different species of bacteria have evolved new transcription factor proteins by duplication, divergence and sometimes subsequent loss of transcription factors. Wagner et al. (2007) have shown that transcription factor binding site abundance is under selection, and varies considerably between species.

Topological changes of real gene networks also occur on very short evolutionary timescales, especially in higher eukaryotes (Stone and Wray, 2001). In contrast to gain and loss of transcription factor proteins, these changes are caused by mutations in promotors that produce novel transcription factor binding sites.

At an even finer grain, Moses et al. (2003) have shown that within a transcription factor binding site (TFBS), some nucleotide positions show greater variation than other nucleotide positions, between related species and within the same genome. There is evidence that the more degenerate positions (i.e. those with lower information content) in the TFBM are those where the TF does not make much contact with the DNA, i.e. where the total stabilization energy in that column of the TFBM is low (Mirny and Gelfand, 2002).

We simulate gene network evolution to investigate under what conditions TFBMs evolve to improve evolvability.

Methods

Our gene network model considers N interacting units. Each unit consists of a transcription factor binding site (TFBS) that produces a transcription factor with a particular transcription factor binding matrix (TFBM). The interaction between this matrix and the TFBS sequence determines how a transcription factor will bind to the transcription factor binding site. All TFBSs have length K nucleotides. At each position in the TFBS , one of four bases can be present (A, T, G, or C). The TFBM is of size 4 by K. Each entry in the matrix contains a real number between 0 and 1. Each number represents the binding strength contribution of the nucleotide at that position in the TFBS to the total TF binding strength upon that TFBS. For simplicity, our TFBM implementation assumes that the contribution of different positions along the TFBS to the binding strength is independent.

The edges of the gene network graph are directed *from* the gene producing the TF *to* the gene possessing the TFBS. The strength of each edge is calculated as described above, by summing independent binding strength contributions from each position in the TFBS as specified in the TFBM. A low accumulated strength represents weak binding between that TFBM_i and TFBS_j. A larger binding strength represents

tighter binding. To specify a *desired* gene network topology, we define a certain binding strength as required for 'ideal' binding. A binding strength greater than this optimum is considered maladaptive. The biological justification for this is the implicit assumption that the TF should be sensitive to modifiers of binding. If it binds to strongly then the protein that it stimulates might as well have been constitutively expressed. More precisely, the fitness contribution of one edge of the gene network corresponds to the inverse of the Euclidean distance between the individual's topology and the desired topology, as given by:

$$f = \frac{1}{1 + \sqrt{\sum_i \sum_j \left((s_{ij}/\lambda) - t_{ij}\right)^2}} \quad (1)$$

where s_{ij} is the binding strength for connection TFBM$_i$ and TFBS$_j$; t_{ij} is 0 or 1 depending on whether a connection is desired to be present or absent (respectively); and λ is the ideal binding strength ($\lambda = 3$). If the Euclidean distance of a topology from an ideal topology X is less than its Euclidean distance from all other ideal topologies, then it is classified as topology type X. An ideal topology is one where the above fitness score is maximized. Note we do not consider the dynamics of the gene network in generating its topology, e.g. we assume TFs are always produced constitutively. Our selection function acts directly on the non-functional topology of the gene network.

Environments are modeled simply as desired gene network topologies. For example in environment A, individuals are optimal that have topology T_a. In a different environment, B, an altogether different topology might be required to survive, T_b.

The parameters of each gene network (i.e. TFBS and TFBM for each gene) are evolved using a microbial genetic algorithm (Harvey, 2001). There are NK discrete and $4NK$ real-valued parameters forming the genotype of each individual. The winner of a randomly chosen pairwise tournament replaces some of the loser's genes with its own. Each TFBM or TFBS of the loser is replaced by the corresponding genome parts of the winner with 0.9 probability. The resulting genome then undergoes mutation. Each TFBM or TFBS will mutate with $1/NK$ probability. On average, one of the full set of TFBSs changes to a base chosen at random. Mutation in the TFBM is implemented as a random displacement on every binding strength drawn uniformly from a Gaussian distribution with mean 0 and variance 0.01. Each strength in the binding matrix is forced to stay within the range 0 to 1. When a mutation takes it out of this range, it is reflected back. A generation is defined as P microbial tournaments, where P is the size of the population. All evolutionary experiments were conducted with populations of 100 individuals. We have not investigated how the findings depend on the choice of genetic algorithm. The major point to note is that TFBMs evolve much more slowly that TFBSs.

Several previously defined measures of robustness and evolvability are considered (Ciliberti et al., 2007; Zhu and Freeland, 2006). *Mutational robustness* is the fraction of one-mutant neighbours of the genome that are also viable. In this case, viability means that the gene network topology resembles the desired topology more closely than all other topologies (with the same number of genes). This is a measure of neutrality or redundancy of coding. *Topology connectivity* is the number of distinct 1-mutant topologies reachable from any one topology. High connectivity implies easy conversion between topologies.

The above properties can be calculated by studying the individual's *hierarchical metagraph* (see Figure 1). In this graph, a higher level node (large circle) represents a particular gene network topology (phenotype). Higher level nodes are connected if by one TFBS mutation, topology A can become topology B, and vice versa. Within a higher level node of a metagraph, there is another embedded graph whose nodes are the genome sequences of TFBSs that can sustain the particular gene network topology represented by the higher level node (i.e. sequence nodes). The higher level node can alternatively be thought of as a labeling of lower level sequence nodes. We define the *connectivity matrix* of a metagraph as the number of one-mutant neighbours connecting topology t_i to topology t_j, over all possible topology transitions.

It is clear that the hierarchical metagraph will depend critically on the evolved TFBMs. The TFBMs *shape* the phenotypic effect of a TFBS mutation. We define the *connectivity variance* simply as the variance over all values of the connectivity matrix. We use this measure as an index of the navigability of the metagraph. The measure synthesizes the system's mutational robustness and its topological connectivity. Our hypothesis is that evolvability in variable environments increases as the connectivity variance decreases, improving the navigability of gene network topology space, as promoter sequence space is explored. Expressed in another way, we propose that TFBM evolution structures the exploration distribution in phenotype space, by reducing connectivity variance.

Let us illustrate what a hierarchical metagraph is and how we can measure its navigability. We will use the simplest gene transcription network: a one-gene network (see Figure 1). There are only two possible topologies: self-binding (t_1) or non self-binding (t_2). Let's suppose that the length of the promoter region is 1. Therefore, there are only 4 possible sequences: A, G, C, or T. The self-binding would depend on the transcription factor binding strength. There are three possible scenarios that can arise. In the first scenario, all possible sequences generate topology t_1 and no sequences result in topology t_2. There are 12 (i.e. BNK) ways in which a sequence can mutate from sequence S_i to S_j where S_j is only one mutation away. In the first case, all of the mutations are neutral. There are 12 ways of going from t_1 to

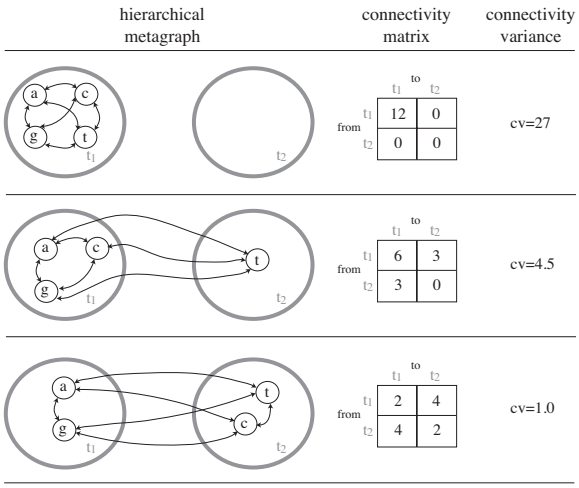

Figure 1: Hierarchical metagraph and the notions of its *connectivity matrix* and *connectivity variance* in the simplest possible one-gene regulatory network scenario.

t_1, and 0 ways of going from t_1 to t_2, t_2 to t_1, t_2 to t_2. The connectivity variance is the highest possible, in this case 27. In the second scenario, there is one sequence that produces t_2. The number of one-neighbour mutants shifts accordingly and the connectivity variance drops. In the final case, the connectivity variance is minimal.

From the example illustrated in Figure 1, it should be clear that the connectivity variance will be lower for more easily navigable metagraphs. Accordingly, easily navigable metagraphs will reduce the time required to adapt to any topology, including novel topologies. Connectivity variance is an index of evolvability in variable environments. We demonstrate that the properties of the metagraph, as measured by its connectivity variance, promote evolvability in this selection scenario.

Results

We first present an experiment where a population of 3-node gene networks are evolved in two different environments. The length of each TFBS, K, is 5. In one environment the target topology is a feedforward loop and in the other environment it is a feedback loop (see Figure 2A, the difference is colored red). The evolved sequences and TFBMs for the best individual are shown in Figures 2B and 2C, respectively. The result is a set of TFBMs such that, with a total of 3 site mutations (colored red) in each of the relevant TFBSs, the desired change in topology is achieved. Once the environmental transitions have been experienced several times, adaptation occurs without significant changes to the TFBMs.

Figure 2D shows that the time taken to adapt to a particular environment decreases over evolutionary time. Each environment was presented for 1000 generations. While it

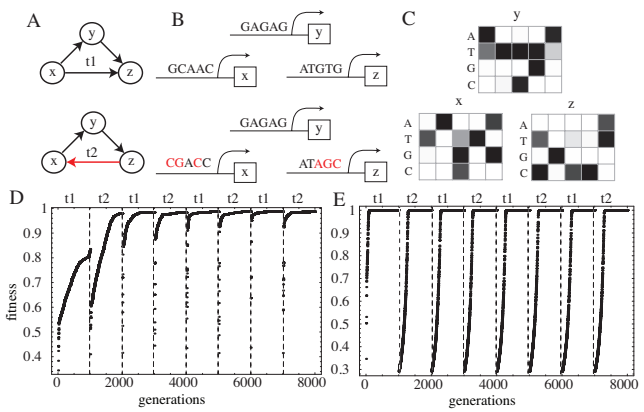

Figure 2: Example evolutionary run with variable environments for a 3-node GRN. [A] Two topologies evolved for: feedforward (t_1) and feedback loop (t_2). [B] Evolved promoter regions for each topology (TFBSs). [C] Evolved set of TFBMs. [D] Fitness versus Generations. Dashed vertical lines represent changes of environment. [E] Control experiment with direct encoding of connectivity shows no evolution of evolvability.

takes the population around 800 generations to adapt to environment B the first time it is encountered, this time drops to around 250 generations on the second occasion. From the third occasion onwards, the adaptation time to environment B is under 100 generations. The time to adaptation reaches steady-state after several environmental transitions. This was also observed for runs using different sized networks, lengths, and numbers of environments. For comparison, Figure 2E shows that when a direct encoding of the gene network is used, i.e. a binary connectivity matrix, there is no improvement in time to adaptation.

For a given size of gene network N, there is a minimum TFBS length K that is required to evolve perfect fitness in W distinct environments, see Figure 3 . This is shown in experiments conducted with 2- and 3-node gene networks with different numbers of variable environments (W) and different lengths of promoter sequences (K). Each data point in Figure 3 is an average over 100 evolutionary runs. For an N sized network, $2^{(N^2)}$ different topologies are possible. For any given experiment, the W different topologies the population would be evolved for were chosen at random from all possible topologies. All experiments ran for the same number of generations, 160000. The number of generations per transition is the same for all experiments, 1000. In Figure 3 we show the proportion of populations in which each agent fully adapts (top) and the time to adaptation (i.e. the time taken for the best individual in the population to reach 0.95 of optimal fitness) (bottom) as a function of W and K. Each point in the surface corresponds to the mean over only those populations where each individual adapted to all W environ-

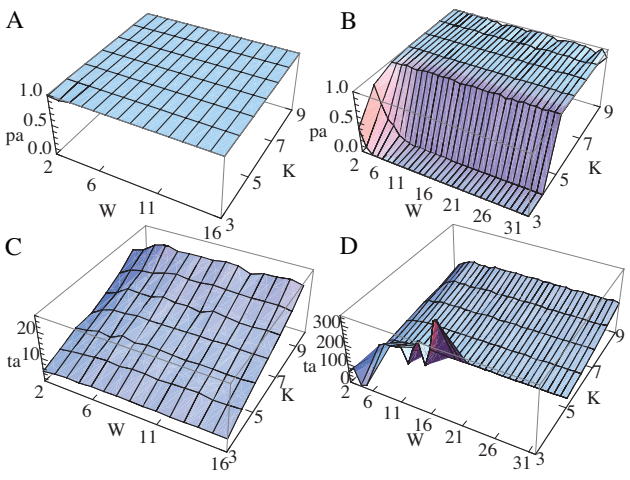

Figure 3: Top: The proportion of 100 independent populations that adapted to all W environments during the last round of environmental presentations as a function of the number of environments presented during evolution (W) and the length of the promoter sequence (K), for $N = 2$ [A] and $N = 3$ [B]. Bottom: The mean time to adaptation of the best individual in the population during the last round of presentations over all environments as a function of W and K, for $N = 2$ [C] and $N = 3$ [D].

ments. No points are shown in conditions where this was not the case.

Referring to Figure 3 C, for $N = 2$ gene networks, TFBS lengths of $K = 3$ are capable of evolving to adapt to all 16 possible topologies. With larger Ks, the adaptation time increases, presumably because the space of possible TFBS sequences also increases. Thus, finding the appropriate sequence that creates the desired topology becomes harder. Interestingly, while the search space becomes larger exponentially with K as given by 4^{NK}, the time to adaption increases only linearly.

Referring to Figure 3 D ,for 3-node GRNs there is a lower limit to K below which adaptation to all environments is not possible. For the minimum, $K = 3$, populations can adapt reliably, only to fixed environments $W = 1$. When tested in variable environments, gene networks with $K = 3$ fail to adapt to 0.95 optimal fitness before the environmental transition at 1000 generations. For $K = 4$ the situation is improved. Populations can adapt to up to 8 different environments. For higher W, the population again starts failing to adapt in time. For $K > 4$, the populations can adapt to varying environments for all W conditions tested.

Does evolution under variable environments increase evolvability? In order to address this question, we tested the ability for evolved individuals to rapidly adapt to *novel* envi-

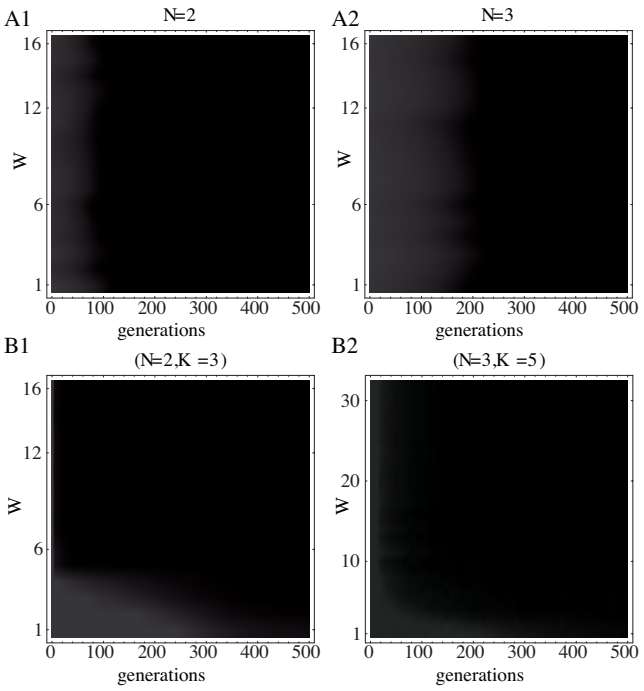

Figure 4: Evolvability. $Generations$ (x-axis) versus environmental variability W. Shades of gray represent the mean of the best individual's fitness over 100 independent evolutionary runs. Black represents poorly adapted populations. White represents well adapted ones. Four conditions are shown: smallest (top row) and largest (bottom row) Ks that evolved successfully under all W conditions tested, for networks of size 2 (left column) and 3 (right column).

ronments. We seeded a new population with genetic 'clones' from the TFBMs of the best evolved individual after 160000 generations. The TFBSs were chosen at random for each individual in the population. Taking into account the topologies the original population had previously evolved for, we set the new population to evolve to achieve a topology it had not previously been exposed to. Different populations were seeded with TFBMs evolved under several conditions. For each condition, the experiment was repeated 100 times.

In Figure 4 (bottom) we show the population's ability to evolve to novel environments under two conditions: $N = 2$, $K = 3$ and $N = 3$, $K = 5$. For all fixed environment conditions $W = 1$, the mean time to adaptation is much larger than for individuals evolved under variable environments $W > 1$. In fact, the ability to rapidly adapt to a novel environment improves with the number of environments that the population was previously evolved for. Finally, all of the conditions show a ceiling effect for sufficiently-varied environments, after which the time to adaptation ceases to improve. For comparison, Figure 4 (top) shows that a direct encoding does not have this property of improved evolvabil-

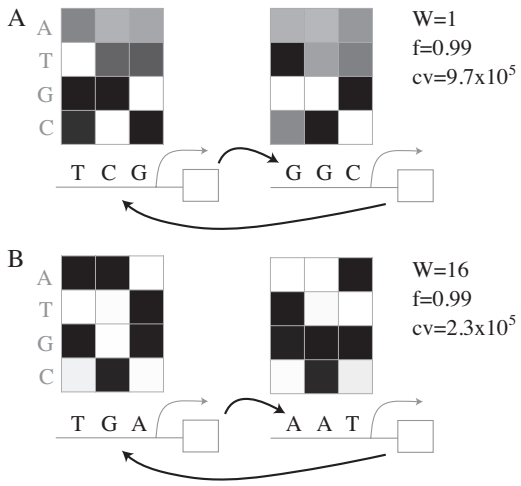

Figure 5: TFBMs evolved for the same topology under fixed [A] and variable [B] environments. Fitness for the feedback topology is equally good for both TFBMs (0.99). However, the connectivity variance for individual B is lower than the connectivity variance for individual A.

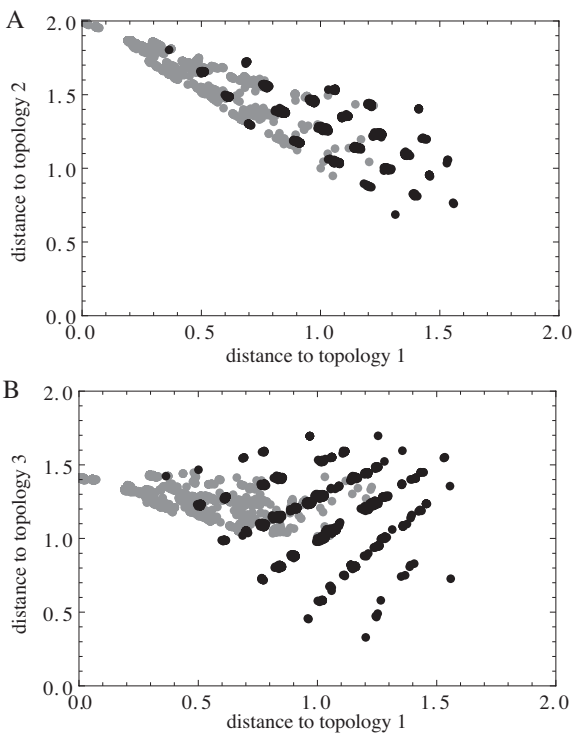

Figure 6: The exploration distribution for the best individual ($N = 2, K = 3$) evolved for a small ($W = 2$, gray) and large ($W = 15$, black) range of environments. The best individual evolved for topology one, t_1, is mutated 5000 times. On the x- and y-axes are shown the Euclidean distance of the resulting mutants to the topologies t_1, t_2 (top) and t_1, t_3 (bottom). Both individuals were evolved previously for t_1, t_2. However, t_3 (a fully connected topology) was evolved for in either case, i.e. it is novel.

ity to a novel environment, given a history of evolution with previously variable environments.

Greater variability of environments experienced during evolution improves evolvability according to the above measure. But why are TFBMs evolved for variable environments more evolvable to new environments than TFBMs optimised for a fixed environment? While two individual's TFBMs can be equally fit for environment T_x, the TFBMs from the individual that has evolved to change to different topologies generates a different exploration distribution to the TFBM evolved in stationary environments.

Figure 5 shows a case where two individuals (A and B) are equally fit for environment T_x, but with different TFBMs. A has evolved in fixed environments while B has evolved for variable environments. While both solutions are equally well adapted to produce the feedback loop topology (i.e. both have fitness of 0.99), the connectivity variance of the gene network evolved for variable environments (B) is less than a quarter of the connectivity variance of the gene network evolved for fixed environments (A). The implication is that the TFBMs evolved for individual B shape the TFBS fitness landscape such that all other topologies are more easily reachable from the present topology. This is not the case for TFBMs evolved for individual A in a static environment.

The exploration distribution can be directly visualized in Figure 6. Individuals evolved in high variability environments, $W = 15$, show a more diffuse exploration distribution to those evolved in low variability environments, $W = 2$. Individuals evolved for $W = 2$ show a tighter exploration distribution, passing from t_1 directly to t_2 without approaching close to other topologies.

Does the connectivity variance always decrease with the number of environments that the populations are evolved for? In order to answer this, the connectivity variance was calculated for all successfully evolved TFBMs for 2- and 3-node networks (see Figure 7). For the 2-node case, the study was exhaustive: taking into consideration all possible sequence configurations ($4^{(}N*K)$) and all possible topologies ($2^{(}N * N)$). For the 3-node case a sample of 1000 different sequence configurations was chosen at random.

The connectivity variance drops as a function of W, as hypothesised (see Figure 7). The connectivity variance results also explain the evolvability ceiling effect observed in Figure 4, where after a certain number of variable environments ($W > 5$) the ability to rapidly adapt to novel environments ceases to improve. The connectivity variance reaches its minimum also for $W > 5$.

Finally, we can compare the evolution of the connectivity variance for the two extreme conditions: fixed or variable

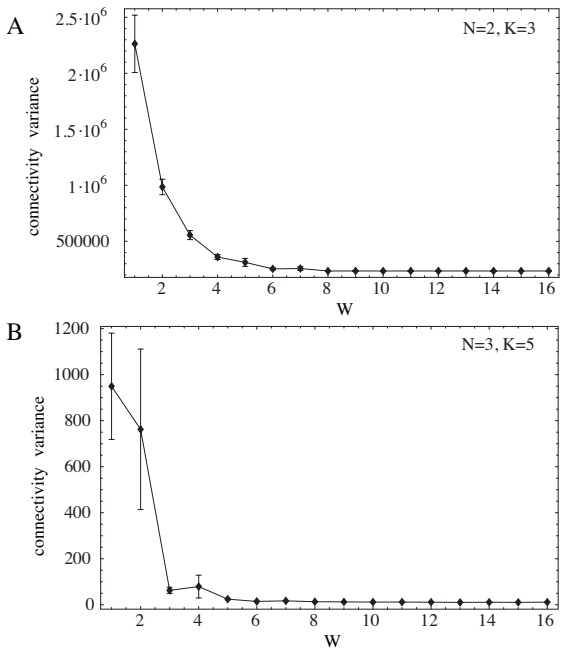

Figure 7: Connectivity variance as a function of variable environments (W). Two conditions are shown: [A] for $N = 2$ and $K = 3$, and [B] for $N = 3$ and $K = 5$. Each point represents the mean over 100 runs (bars represent the standard error).

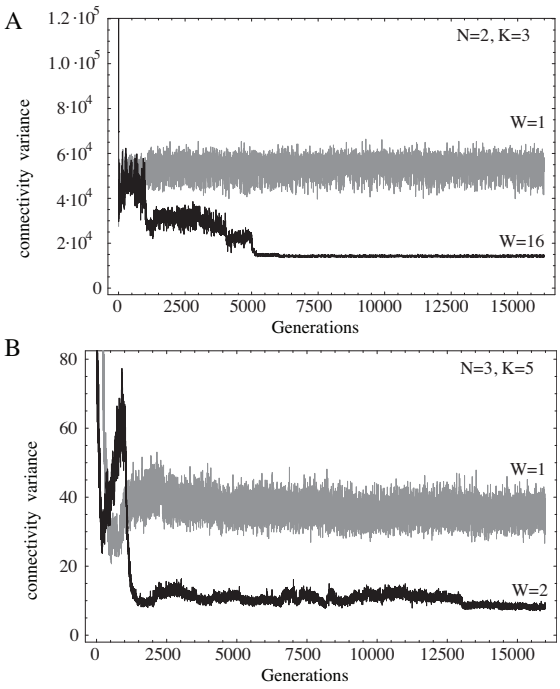

Figure 8: The evolution of evolvability. Connectivity variance of the best individual in the population over evolutionary time. Examples for fixed (gray) and variable (black) environments are shown for 2-node [A] and 3-node [B] GRNs.

environments. In Figure 8 we show the connectivity variance calculated from the TFBMs of the best individual in the population at every generation in four different conditions: fixed (gray) and variable (black) environments are shown for 2-node [A] and 3-node [B] GRNs. Under all conditions the populations evolved successfully (not shown). As can be appreciated in Figure 8, the connectivity variance of populations evolving in varying environments tends to evolve towards lower values (and thus higher evolvability), compared to the connectivity variance of the population evolving in fixed environments. This demonstrates the evolution of evolvability under variable environments.

Discussion

We gave an example of the evolution of evolvability in a system undergoing natural selection in variable environments. In our model, evolvability arose from the ability of the TFBM to evolve to shape the exploration distribution resulting from TFBS mutations. When populations were evolved in variable environments, the TFBMs allowed improved navigability in TFBS space as described by the connectivity variance measure.

The model lacks many features of real GRNs. In reality, promotor sequences are much longer than TFBSs, enhancer proteins modify TF binding, the expression of downstream TFs may be dependent on upstream TFs, and fitness depends on the dynamics of the GRN rather than on its topology alone. Introducing such features whilst maintaining TFBM-TFBS interactions is a challenge, and is likely to uncover further adaptations that could lead to unlimited heredity of exploration distribution variations.

Acknowledgements

This work was supported by the ESIGNET 6th Framework EU grant. Thanks to Richard Goldstein, Jon Rowe and Dov Stekel for helpful discussions.

References

Altenberg, L. (1994a). The evolution of evolvability in genetic programming. In Kinnear, K., editor, *Advances in genetic programming*, pages 47–74. Cambridge, MA: MIT Press.

Altenberg, L. (1994b). Evolving better representations through selective genome growth. *Proc. of the First IEEE Conf. on Evolutionary Computation*, 1:183–187.

Arthur, W. (2004). The effect of development on the direction of evolution:toward a twenty-first century consensus. *Evolution and Development*, 6:282–288.

Babu, M. M., Teichmann, S. A., and Aravind, L. (2006). Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol*, 358(2):614–33.

Barnett, L. (2003). *Evolutionary Search on Fitness Landscapes with Neutral Networks*. PhD thesis, School of Cognitive and Computer Science, University of Sussex, Falmer, Brighton.

Berg, J., Willmann, S., and Lässig, M. (2004). Adaptive evolution of transcription factor binding sites. *BMC Evol Biol*, 4(1):42.

Bloom, J., Labthavikul, Sy, T., Otey, C., and Arnold, F. (2006). Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. USA*, 103(15):5869–5874.

Ciliberti, S., Martin, O., and Andreas, W. (2007). Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Computational Biology*, 3(2):0164–0173.

Collado-Vides, J., Magasanik, B., and Gralla, J. (1991). Control site location and transcriptional regulation in escherichia coli. *Microbiol Rev.*, 55(5):371–394.

Conrad, M. (1990). The geometry of evolution. *BioSystems*, 24(61-81).

Earl, D. and Deem, M. (2004). Evolvability is a selectable trait. *Proc. Natl. Acad. Sci. USA*, 101(32):11531–11536.

Force, A., Cresko, W., Pickett, F., Proulx, S., Amemiya, C., and Lynch, M. (2005). The origin of subfunctions and modular gene regulation. *Genetics*, 170:433–446.

Gelfend, M. (2006). Evolution of transcriptional regulatory networks in microbial genomes. *Current Opinion in Structural Biology*, 16:420–429.

Gruau, F., Whitley, D., and Pyeatt, L. (1996). A comparison between cellular encoding and direct encoding for genetic neural networks. In Koza, J. R., Goldberg, D. E., Fogel, D. B., and Riolo, R. L., editors, *Genetic Programming 1996: Proc. of the First Annual Conference*, pages 81–89, Stanford University, CA, USA. MIT Press.

Hansen, T. (2006). The evolution of genetic architecture. *Annu. Rev. Ecol. Evol. Syst.*, 37:123–157.

Harvey, I. (2001). Artificial evolution: a continuing SAGA. In Gomi, T., editor, *Evolutionary Robotics: From Intelligent Robots to Artificial Life.* Springer-Verlag LNCS 2217.

Hornby, G. and Pollack, J. (2002). Creating high-level components with a generative representation for body-brain evolution. *Artificial Life*, 8:223–246.

Kashtan, N. and Alon, U. (2005). Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci USA*, 102(39):13773–8.

Kirchner, M. and Gerhart, J. (1998). Evolvability. *Proc. Natl. Acad. Sci. USA*, 95:8420–8427.

Kussell, E. and Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743):2075–8.

Lipson, H., Pollack, J., and Suh, N. (2002). On the origin of modular variation. *Evolution*, 56:1549–1556.

Matic, I., Radman, M., Taddei, F., Picard, B., Boit, C., Bingen, E., Denamur, E., and Elion, J. (1997). Highly variable mutation rates in commensal and pathogenic e.coli. *Science*, 277:1833–1834.

McGregor, S. and Fernando, C. (2005). Levels of description: A novel approach to dynamical hierarchies. *Journal of Artificial Life*, 11:459–472.

Mirny, L. and Gelfand, M. (2002). Structural analysis of conserved base pairs in protein-dna complexes. *Nucleic Acids Res*, 30(1704-1711).

Moses, A. M., Chiang, D. Y., Kellis, M., Lander, E. S., and Eisen, M. B. (2003). Position specific variation in the rate of evolution in transcription factor binding sites. *BMC Evol Biol*, 3:19.

Prescott, D. and Rozenberg, G. (2002). How ciliates manipulate their own dna - a splendid example of natural computing. *Natural Computing*, 1:165–183.

Rando, O. and Verstrepen, K. (2007). Timescales of genetic and epigenetic inheritance. *Cell*, 128:655–668.

Reisinger, J. and Miikkulainen, R. (2007). Acquiring evolvability through adaptive representations. In *Proc. of Genetic and Evolutionary Computation Conference*, pages 1045–1052.

Smith, T., Husbands, P., Layzell, P., and O'Shea, M. (2002). Fitness landscapes and evolvability. *Evolutionary Computation*, 1(10):1–34.

Stone, J. and Wray, G. (2001). Rapid evolution of cis-regulatory sequences via local point mutations. *Mol. Biol. Evol.*, 18(9):1764–1770.

Stormo, G. (2000). Dna binding sites: representation and discovery. *Bioinformatics*, 16(1):16–23.

Tenaillon, O., Taddei, F., Radman, M., and Matic, I. (2001). Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. *Res. Microbiol*, 152:11–16.

Tirosh, I., Weinberger, A., Bezalel, D., Kaganovich, M., and Barkai, N. (2008). On the relation between promoter divergence and gene expression evolution. *Molecular Systems Biology*, 4(159):1–11.

Toussaint, M. (2003). *The evolution of genetic representations and modular adaptation*. PhD thesis, Institute for Neuroinformatics, Ruhr University of Bochum, ND 04, 44780 Bochum—Germany.

Turney, P. (1987). Increasing evolvability considered as a large-scale trend in evolution. In Wu, A., editor, *Proc. of 1999 Genetic and Evolutionary Computation Conference Workshop Program*, pages 43–46.

Volkert, L. and Conrad, M. (1998). The role of weak interactions in biological systems: the dual dynamics model. *J. Theor. Biol.*, 193:287–306.

Wagner, G. and Altenberg, L. (1996). Complex adaptations and evolution of evolvability. *Evolution*, 50:329–47.

Wagner, G., Otto, W., Lynch, V., and Stadler, P. (2007). A stochastic model for the evolution of transcription factor binding site abundance. *J Theor Biol*, 247(3):544–553.

Zhu, W. and Freeland, S. (2006). The standard genetic code enhances adaptive evolution of proteins. *Journal of Theoretical Biology*, 239:63–70.