CENTRE DE RECERCA MATEMÀTICA

This is a preprint of: *The topology of robustness and evolvability in evolutionary systems with genotype–phenotype map*
Journal Information: *Journal of Theoretical Biology*,
Author(s): E. Ibanez-Marcelo and T. Alarcon.
Volume, pages: 356 1-19, DOI:[doi.org/10.1016/j.jtbi.2014.04.014]

The topology of robustness and evolvability in evolutionary systems with genotype–phenotype map

Esther Ibáñez-Marcelo ^{a,b,*}, Tomás Alarcón ^{a,c}

^a Centre de Recerca Matemàtica, Edifici C, Campus de Bellaterra, 08193 Bellaterra, Barcelona, Spain

^b Departament de Matemàtica Aplicada I, Universitat Politècnica de Catalunya, 08028 Barcelona, Spain

^c Departament de Matemàtiques, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

HIGHLIGHTS

- Characterise robustness and evolvability using genotype–phenotype networks topology.
- We show that phenotype network exhibits small-world property.
- Evolvability is a robust property respect to attack of randomly selected genotypes.
- Targeted attacks are much more efficient in breaking the giant component.
- Results support that robustness facilitates innovation.

ARTICLE INFO

Article history:

Received 3 October 2013

Received in revised form

26 March 2014

Accepted 14 April 2014

Available online 30 April 2014

Keywords:

Complex networks

Genetic Regulatory Networks (GRN)

Modeling

Percolation

ABSTRACT

In this paper we formulate a topological definition of the concepts of robustness and evolvability. We start our investigation by formulating a multiscale model of the evolutionary dynamics of a population of cells. Our cells are characterised by a genotype–phenotype map: their chances of survival under selective pressure are determined by their phenotypes, whereas the latter are determined their genotypes. According to our multiscale dynamics, the population dynamics generates the evolution of a genotype–phenotype network. Our representation of the genotype–phenotype network is similar to previously described ones, but has a novel element, namely, our network contains two types of nodes: genotype and phenotype nodes. This network representation allows us to characterise robustness and evolvability in terms of its topological properties: phenotypic robustness by means of the clustering coefficient of the phenotype nodes, and evolvability as the emergence of giant connected component which allows navigation between phenotypes. This topological definition of evolvability allows us to characterise the so-called *robustness of evolvability*, which is defined in terms of the robustness against attack (i.e. edge removal) of the giant connected component. An investigation of the factors that affect the robustness of evolvability shows that phenotypic robustness and the cryptic genetic variation are key to the integrity of the ability to innovate. These results fit within the framework of a number of models which point out that robustness favours rather than hindering evolvability. We further show that the corresponding phenotype network, defined as the one-component projection of the whole genotype–phenotype network, exhibits the small-world phenomenon, which implies that in this type of evolutionary system the rate of adaptability is enhanced.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Understanding how complex global behavioural traits (phenotypes) emerge from the interactions between individual genes and their products is one of the major challenges of modern biology. Phenotypes arise from networks of interactions between genes and gene products, which ultimately regulate gene expression.

These networks of gene regulation are dynamical systems whose complexity partially stems from the fact that they are non-linear, high-dimensional dynamical systems, also and foremost, because they are shaped up by evolution by natural selection. Natural selection acts upon gene regulatory networks (GRNs) so that they evolve to exhibit properties such as robustness (i.e. resilience of the phenotype against genetic alterations, Wagner, 2007), canalisation (i.e. the ability for phenotypes to increase their robustness as time progresses, Waddington, 1942a,b) or non-uniqueness (i.e. different genotypes leading to the same phenotype, Ciliberti et al., 2007a,b). Moreover, they are under the influence of noise of both

* Corresponding author.

E-mail address: eibanez@crm.cat (E. Ibáñez-Marcelo).

internal (molecular noise in the regulatory system itself) and external (unpredictable changes in the environment) characters.

In order to study the properties and issues regarding the genotype–phenotype maps, several models have been studied (Wagner, 2012): RNA, circuits of gene regulation and metabolic networks. Concerning RNA molecules as model of the genotype–phenotype map (Fontana and Schuster, 1998), the *genotype* of each RNA molecule consists of a sequence of nucleotides. There are four such nucleotides, so for sequences of length L , the size of the genotype space is 4^L . The *phenotype* of the RNA molecule refers to the fold or three-dimensional conformation, which determines the biochemical function of the molecule. The folded structure of an RNA sequence, which is a proxy for its phenotype, but still lies far from defining its function, is determined by the sequence (genotype) in a many-to-one way, i.e. many different genotypes give rise to the same phenotype. Such non-uniqueness has led to the concept of the neutral network, first introduced in Lipman and Wilbur (1991) and Schuster et al. (1994), i.e. a network whose nodes correspond to genotypes, all with the same phenotypes, with edges between those nodes which differ by only one nucleotide (Wagner, 2007). This system has been used extensively in the study of the genotype–phenotype map, in particular, those issues regarding its evolutionary properties, such as the role of phenotypic robustness in evolvability and adaptation (Wagner, 2011, 2012). Recently, the topology of the RNA genotype–phenotype space, composed of an intermingled set of neutral networks, has been analysed (Aguirre et al., 2011).

Gene regulatory networks (GRNs) have also been used to analyse properties of the genotype–phenotype map, in particular several variants of the model were introduced by Wagner to study phenotype plasticity (Wagner, 1996). These models are dynamical systems for the expression levels of the corresponding genes and are characterised by two elements: a matrix whose entries specify the character of the interaction between two genes (usually, activation or inhibition) and, possibly, the intensity of such interaction, and a series of rules for the time evolution of the expression levels of the genes involved. The entries of the corresponding matrix are the *genotype* of the GRN. The *phenotype* is the steady-state gene expression yielded by the dynamics. There are many such genotypes that produce the same phenotype, which allows to extend the concept of neutral network to GRN, where nodes correspond to different matrices (producing the same steady-state) and links exist between nodes if the corresponding matrices differ only in one regulatory interaction. Models of GRNs have been used to study phenotype plasticity (Wagner, 1996), robustness and innovation in circuits of gene regulation (Ciliberti et al., 2007a,b), and canalisation (Siegal and Bergman, 2002; Bergman and Siegal, 2003), among other issues.

Metabolic networks are the third class of systems that have been used to assess properties regarding robustness and innovation (Ndifon et al., 2009; Rodrigues and Wagner, 2009; Samal et al., 2010; Rodrigues and Wagner, 2011). They are formed by thousands of enzyme-catalysed chemical reactions. These networks are responsible for supplying cells with energy (i.e. ATP) and the molecular building blocks cells need to grow. The *genotype* space for this system consists of the space of all the possible metabolic networks, whereas the *phenotype* corresponds to the secondary metabolites the metabolic network is able to synthesise, the molecules they can use as energy sources, the ability to detoxify certain waste products, etc. (Wagner (2012)). Innovations in these aspects do not always appear as the result of gene mutations that give rise to new enzymes. They can also arise through novel combinations and utilisation of existing elements.

Although the emphasis of this work is on evolutionary properties of systems endowed with a genotype–phenotype map, the utilisation of existing elements is a major concern in systems biology, since understanding how phenotypes arise from cells

genetic makeup and the corresponding networks of gene regulation is a major issue in several important areas from synthetic biology to therapeutics, with important progress being made in this area. For example, Karr et al. (2012) have proposed a whole-cell computational model that predicts phenotype from genotype for bacterium *Mycoplasma genitalium*.

The genotype–phenotype map exhibits a number of emergent properties, the study of which is the main aim of the present paper. Two important concepts arising when studying the genotype–phenotype map are robustness and evolvability which correspond to the resilience to mutations and the ability for adaptation, respectively (Wagner, 2007, 2008a,b). As a working definition of robustness, one could say that a biological system is robust if it continues to function subject to perturbations. The perturbations can be genetic, e.g. mutations, or non-genetic, e.g. an environmental change. As far as the present work is concerned, we consider only mutational robustness which we take to be the ability to retain the same phenotype upon occurrence of genetic mutations.

Evolutionary adaptation of a population to, say, a new environment requires evolutionary innovation or evolvability, namely, new, better adapted phenotypes must arise within the population. Such innovations are achieved over many generations by means of genetic mutations. Most of these mutations are known to be either detrimental or neutral (Eyre-Walker and Keightley, 1990; Sawyer et al., 2007; Soskine and Tawfik, 2010). However, rarely, one or several mutations produce new phenotypes that are better adapted to the new conditions. These rare mutations are the drivers of Darwinian evolution. We consider evolvability as defined by Wagner (2008b), i.e. we say that a system is evolvable if mutations in it can produce heritable phenotypic variation.

Intuitively, robustness and evolvability seem to be in conflict, since, by definition, robust phenotypes are resilient to the effects of mutations. However, there is mounting evidence which hints otherwise, i.e. phenotypic robustness facilitates evolutionary innovation (Wagner, 2011, 2012). According to this view, the genotype space has two generic properties which allow to reconcile robustness and evolvability. The first of these properties is the existence of neutral networks, i.e. large connected sub-networks of the space of genotypes which can be navigated in small, mutation-induced steps with no change in phenotype. The second of these features regards the so-called *genotype neighbourhood*, i.e. the set of genotypes accessible from any one genotype in a prescribed number of mutations. A simple measure of how phenotypically variable is a genotype is the number of different phenotypes reached in a genotype neighbourhood, i.e. how many different phenotypes are easily accessible by mutation(s) (Wagner, 2008a). The first of these properties allows preservation of the phenotype with changing genotypes, thus creating a substantial amount of divergence within a population, whereas the second property, neighbourhoods, allows genotypes within a neutral network to explore different phenotypes.

The argument to support that robustness favours evolvability rather than hindering it can be summarised as follows (Wagner, 2011, 2012): large neutral networks containing many genotypes with the same phenotype have, collectively, a much larger neighbourhood, i.e. corresponding to the (disjoint) union of all the neighbourhoods of all the nodes within the neutral network, than non-robust genotypes, which will be typically isolated in genotype space.

These issues regarding the relation between robustness and evolvability may seem to be of a rather theoretical character. However, they are fundamental to understanding the evolutionary dynamics of systems such as cancer (Kitano, 2004; Tian et al., 2011). Kitano (2004) points out that cancers can be seen as robust systems with respect to its proliferative potential, which is maintained in the presence of several anti-cancer therapies and environmental factors, such as the effect of the immune system. Kitano's thesis is that, as it occurs in systems engineering, the

development of robustness with respect to some properties leads to fragility with respect to others which should be exploited as therapeutic targets. Whereas this is a valid and useful point of view, it may fall short of a full picture, since it fails to take into account that tumours are also highly evolvable (Tian et al., 2011). According to results recently reviewed by Tian et al. (2011), the evolutionary dynamics of tumours is such that evolvability is greatly enhanced. These two results can be unified under the above paradigm that robustness favours evolvability which implies that, to achieve effective control (therapeutic) strategies, both robustness and evolvability must be tackled (Tian et al., 2011).

The aim of this paper is formulating a simple model of the genotype–phenotype map that assists us in determining the basic elements allowing for coevolution of properties such as robustness and evolvability, as well as feasible control strategies. In particular, we give a topological characterisation of evolvability that allows us to analyse whether evolvability is itself a robust property of our system as well as studying the adaptability properties of our system. Our model is based on the one proposed by Wagner (1996) and variations therein proposed by Ciliberti et al. (2007a,b). A similar model has been used in Siegal and Bergman (2002) and Bergman and Siegal (2003) to model canalisation.

This paper is organised as follows. Section 2 is devoted to a detailed description of our model, including our definition of the genotype–phenotype network. In Sections 3 and 4 we give definitions of the concepts of robustness and evolvability in terms of topological properties of the genotype–phenotype network. In Section 5, we explore whether evolvability itself is a robust property by exploiting its topological characterisation provided in Section 4, as well as issues regarding the ability for adaptive behaviour. Last, in Section 6, we discuss our results and present our conclusions and future avenues for further research.

2. Model formulation

We consider a population of cells whose dynamics is described by a multiscale model (Alarcón et al., 2005) composed of three mutually coupled levels: the microscale, where the intracellular dynamics is modelled in terms of a gene regulatory network (GRN), the mesoscale, i.e. the population dynamics of the population where birth and death rates are assigned according to the phenotype of each cell, and the macroscale, consisting of a model of the evolutionary dynamics of the genotype–phenotype space (to be defined later).

Each of these levels has associated different characteristic time scales which become longer as we move up the hierarchy of the model: from the microscale, where the characteristic time scales correspond to the ones associated to intracellular processes (ranging from seconds to hours), to the mesoscale, characterised by the time scales of the order of the life-time of a cell (ranging from a few days to months or even years), and to the macroscale, where the time scales are those characteristic of evolutionary processes and are measured in generations (Alarcón et al., 2005) (see Fig. 1).

We consider a scenario where selective pressures act on phenotypes. We therefore formulate a model with separation of time scales where we allow individuals to complete the developmental plan encoded in their genotype (to be defined in Section 2.1), determine the corresponding phenotype, and then decide whether it is fit to survive and thrive or not. A similar selection procedure has been used, for example, by Siegal and Bergman (2002) and Bergman and Siegal (2003).

2.1. Microscale: intracellular dynamics of the GRN

Our model considers a population of cells characterised by a pair $\mathcal{G} = (G, g(0))$. G is a matrix accounting for the interaction

between genes (i.e. how a gene product affects the (in)activation of all other genes). The vector $g(0)$ corresponds to the initial pattern of gene expression, which can be interpreted as heritable genetic information: the components of $g(0) = (g_i(0))$ are the states (active or inactive) of each gene when new cells are born. This pair is referred to as the *genotype* in the remaining of this manuscript.

The aim of the microscale dynamics is to provide our model with a representation of the genotype–phenotype map, i.e., a correspondence between $(G, g(0))$ and the phenotype, ϕ . Our model is closely related to the model used by Wagner to study plasticity (Wagner, 1996) and the model used by Bergman and Siegal to study canalisation (Siegal and Bergman, 2002; Bergman and Siegal, 2003). It is also based on the idea, proposed by Stuart Kauffman, that GRNs are dynamical systems and phenotypes or differentiated states correspond to the stable attractors of these dynamical systems (Kauffman, 1993). Within this framework, we represent GRNs as graphs, where genes are the nodes and links represent a regulatory interaction between the genes corresponding to the nodes at either end of the link. This graph can be represented by a matrix A , the so-called adjacency matrix, where $\dim(A) = N_G \times N_G$, with N_G equal to the number of genes. A is defined so that its entries, a_{ij} , are such that $a_{ij} = 1$ if there exists a link between nodes (genes) i and j , and $a_{ij} = 0$, otherwise.

Each gene, labelled by $i = 1, \dots, N_G$, is endowed with a state variable $g_i(t)$, which, at time t , can take two values: $g_i(t) = 1$ if the gene is being expressed and its protein product being synthesised, and $g_i(t) = -1$, otherwise. The matrix $G = (g_{ij})$ is a weighted version of A : $g_{ij} = a_{ij}$ if gene i activates gene j , and $g_{ij} = -a_{ij}$ if gene i inhibits gene j . GRNs are directed graphs and, therefore, both A and G need not be symmetric.

The phenotype is defined as the steady-state of the dynamical system defined by the following set of rules:

1. At $t=0$, that is, at the time of birth of each cell, we fix $g(t=0) = g_0$.
2. At each time step, and for each gene in the GRN, we determine the value of the variable $I_i(t)$, defined by

$$I_i = \sum_{j \in \langle j \rangle_i^{\text{in}}} g_j g_j \quad \text{where } \langle j \rangle_i^{\text{in}} \text{ is the set of in-neighbours of } i. \quad (1)$$

3. We determine the value of the state of each gene at step $t+1$ according to

$$\begin{cases} g_i(t+1) = 1 & \text{if } I_i(t) \geq 0 \\ g_i(t+1) = -1 & \text{if } I_i(t) < 0 \end{cases}$$

4. Steps 2 and 3 are repeated until $t = T \gg 1$. The phenotype corresponding to the genotype $\mathcal{G} = (G, g_0)$, $\phi(\mathcal{G})$, is defined by $\phi(\mathcal{G}) = g(T)$

2.1.1. Selective pressure: viability conditions

We formulate conditions to discard genotypes which give rise to *unviable* phenotypes. Our viability conditions are related by those imposed by Siegal and Bergman (2002) and Bergman and Siegal (2003). These authors consider that a gene regulatory network is in steady state if an average measure analogous to a variance is smaller than a (small) threshold value. If the GRN has not reached a steady state after a certain number of iterations, it is discarded as non-viable. The average involved in this steady state criterion is taken over a period of time of length τ , which implies that solutions with periods longer than τ are non-viable. Here, we discard oscillatory solutions if their period is longer than a threshold length, l_p . In the

Fig. 1. Description of multiscale model. Time measures in microscale: intracellular process then ranging from seconds to hours, mesoscale: life-time of a cell ranging from few days, months or even years, macroscale: generations.

case of these oscillatory solutions, our definition of the phenotype is slightly different: $\phi_i(\mathcal{G}) = 0$ if g_i oscillates, and $\phi_i(\mathcal{G}) = g_i(T)$, otherwise. For example, if $\phi(\mathcal{G})$ oscillates like $(1, 1, 1, 1) \rightarrow (1, -1, 1, 1) \rightarrow (1, 1, 1, 1) \rightarrow (1, -1, 1, 1)$, then $\phi(\mathcal{G})$ is defined as $(1, 0, 1, 1)$.

From our definition of viability conditions, we can see that the shorter we choose l_v , the more the GRNs are going to be disregarded as unviable. Therefore smaller the l_v , the more severe the selection pressure.

2.1.2. GRN topology

Finally, we need to model the topology of the GRNs. We use two different models in order to capture two topological properties. A number of protein-interaction and transcription networks have been found to exhibit the small-world phenomenon (von Noort et al., 2004; Yook et al., 2004; Albert, 2005) and, in view of this, we will use the Strogatz–Watts model to generate the graphs underlying GRNs with the small-world property (Watts

and Strogatz, 1998). This model works as follows. Consider a regular lattice with N_G vertices and k edges per vertex such that $E = kN_G/2$. By rewiring each edge at random with probability p , Watts and Strogatz showed that, between the limits of a regular graph ($p=0$) and a random graph ($p=1$), there exists an intermediate regime between these two extremes where there is a transition whereby the rewired graph exhibits the small-world effect, i.e. they simultaneously show high clustering and short-path length (Newman and Watts, 1999).

A second topological property whose effects we are interested in exploring is the scale-free topology shown to be exhibited by GRNs (Babu et al., 2004), which we generate using the preferential attachment model (Barabasi and Albert, 1999). Greenbury et al. (2010) have shown that the scale-free topology leads to increased robustness with respect to GRN with random, Erdos–Renyi topology. Here we will examine the small-world phenomenon and the scale-free topology that affect the evolvability properties of our system.

2.2. Mesoscale: population dynamics

The mesoscale describes the dynamics of the population of cells, which we assume to be described by a multi-type Wright–Fisher model with mutation (Blythe and McKane, 2007). In our model, cell types correspond to genotypes where n_i is the number of cells with genotype $\mathcal{G}_i$. The Wright–Fisher model is an urn model where the total number of cells $N_c = \sum_i n_i$ is kept constant. The population dynamics is as follows:

1. A cell is chosen at random for proliferation. All cells are equally likely to be chosen. Therefore, the probability of a cell with genotype $\mathcal{G}_i$ to be picked up is equal to n_i/N .
2. With probability μ , the chosen cell is subjected to mutation. In our model (see Wagner, 1996), a mutation corresponds to changing the sign of one and only one randomly chosen, non-zero entry of the matrix G : i.e. $G \rightarrow G'$ where $g'_{ij} = -g_{ij}$, where the ij -edge is randomly chosen with probability $1/E$, with $E = \sum_{i,j} a_{ij}$ is the number of edges (interaction between genes) of the GRN, and $g'_{ik} = g_{ik}$ for all $i \neq l$ and $k \neq j$. We check, by running the microscale dynamics, whether $\phi(\mathcal{G}_i)$ is equal to $\phi(\mathcal{G}')$, where $\mathcal{G}' = (G', g_0)$. We have three possibilities, namely, (i) $\phi(\mathcal{G}_i) = \phi(\mathcal{G}')$, (ii) $\phi(\mathcal{G}_i) \neq \phi(\mathcal{G}')$ and $\phi(\mathcal{G}')$ is viable, and, last, (iii) $\phi(\mathcal{G}_i) \neq \phi(\mathcal{G}')$ and $\phi(\mathcal{G}')$ is non-viable.
3. If there is no mutation (with probability $1-\mu$), then $n_i \rightarrow n_i + 1$. If there is a mutation which results in cases (i) or (ii), then $n_i \rightarrow n_i - 1$ and $n_{i'} \rightarrow n_{i'} + 2$, where $n_{i'}$ is the population of cells with genotype $\mathcal{G}'$. In either case, another cell chosen at random is eliminated, so that the total cell population stays constant.
4. If there is a mutation which results in a non-viable genotype (case (iii)), steps 1 and 2 are repeated until a viable genotype is found.

This description of the cell population dynamics shows that our viability conditions act as a selective pressure on the population: those genotypes more prone to yield oscillations of period longer than l_v are negatively selected.

2.3. Macroscale: evolutionary dynamics of the genotype–phenotype map

The macroscale describes the evolutionary dynamics of the genotype–phenotype space. Our representation of this space is similar, but not identical, to the metagraph proposed by Ciliberti et al. (2007a,b), who analysed this space using a graph which extends the concept of neutral network (Wagner, 2007).

The model of Ciliberti et al. (2007a,b) can be summarised as follows. Consider that all the cells in our population share the same inherited developmental plan, g_0 . Then, genotypes $\mathcal{G} = (G, g_0)$ are characterised by the corresponding matrix G alone. In the previous section, we have defined a mutation as a sign change in one of the non-zero entries of G . To properly characterise this in mathematical terms, let us define the distance between two matrices G and G' , $\text{dist}(G, G')$, as

$$\text{dist}(G, G') = \sum_{ij} \frac{|g_{ij} - g'_{ij}|}{2}, \quad (2)$$

i.e. $\text{dist}(G, G')$ is the number of entries in G and G' with differing signs. Furthermore, consider the set of viable genotypes, $\mathcal{G} = (G, g_0)$. Based on this set, we can define a graph $\Gamma = (V_G, E_G)$, where V_G is the set of vertices and E_G is the set of edges, in the following way:

1. V_G is equal to the set of viable genotypes.
2. E_G is defined as follows: for all pair $\mathcal{G}_i, \mathcal{G}_j \in V$ an edge exists linking both nodes if and only if $\text{dist}(G_i, G_j) = 1$.

Ciliberti et al. (2007a,b) characterise phenotypic robustness and evolvability in terms of certain topological properties of this graph, namely, its modularity structure is related to robustness: if genotypes with the same phenotype reside within modules (i.e. tightly interconnected subsets of V_G Newman, 2010), the phenotype is said to be robust as it is highly likely that gene mutations lead to genotypes producing the same phenotype.

We propose here an extension of this graph by considering a pseudo-bipartite graph, $B = (V_G, V_\phi, E_G, E_\phi)$, where we have two types of nodes and two types of edges:

1. V_G is equal to the set of viable genotypes.
2. V_ϕ is equal to the set of viable phenotypes.
3. E_G , the set of genotype–genotype links, is defined as follows: for all pair $\mathcal{G}_i, \mathcal{G}_j \in V_G$ an edge exists linking both nodes if and only if $\text{dist}(G_i, G_j) = 1$.
4. E_ϕ , the set of genotype–phenotype edges, is given by the phenotype–genotype map: an edge exists between ϕ_i and $\mathcal{G}_j$ if and only if $\phi_i = \phi(\mathcal{G}_j)$.

Note that this definition departs from that of a bipartite graph in which we allow edges between nodes of the same type (i.e. genotype–genotype edges).

2.4. Evolutionary dynamics of the genotype–phenotype map

Our aim is to analyse the evolutionary dynamics and the emergent properties of the genotype–phenotype map. We propose the following multiscale evolutionary dynamics model.

Initial condition: At $t=0$ we fix the initial condition for our evolutionary dynamics as follows:

1. An inheritable developmental programme, g_0 , is randomly generated.
2. N_c GRNs are generated. This is a two step process:
 - N_c replicas of an adjacency matrix are generated at random using the Watts–Strogatz model with rewiring probability p .
 - For each of these adjacency matrices, each non-zero entry is given a positive (negative) weight with probability p_+ ($1-p_+$).
3. For each of the N_c genotypes $\mathcal{G}_i = (G_i, g_0)$, $i = 1, \dots, N_c$, we determine the corresponding phenotype $\phi_i = \phi(\mathcal{G}_i)$ by means of the microscale dynamics described in Section 2.1. If non-viable genotypes are encountered, the corresponding GRN is

discarded and another one is generated. This process is repeated until we have N_c viable genotypes and produces the set of genotypes present in the population at time $t=0$, $S_G(t=0) = \{G_i, i = 1, \dots, N_c\}$, and the set of phenotypes present in the population at $t=0$, $S_\phi(t=0) = \{\phi(G_i), i = 1, \dots, N_c\}$. Note that repeated genotypes and phenotypes are included only once in their respective sets.

4. The initial condition for the genotype–phenotype graph, $B_0 = (V_G(t=0), V_\phi(t=0), E_G(t=0), E_\phi(t=0))$, is fixed as follows:

- The set of genotype nodes, $V_G(t=0) = S_G(t=0)$.
- The set of phenotype nodes, $V_\phi(t=0) = S_\phi(t=0)$. Note that $\text{card}(V_\phi) \leq \text{card}(V_G)$.
- The set of genotype–genotype links, $E_G(t=0)$, is defined as follows: for all pair $G_i, G_j \in V_G(t=0)$ an edge exists linking both nodes if and only if $\text{dist}(G_i, G_j) = 1$.
- The set of genotype–phenotype edges, $E_\phi(t=0)$, is determined by the phenotype–genotype map: an edge exists between $\phi_i \in V_\phi(t=0)$ and $G_j \in V_G(t=0)$ if and only if $\phi_i = \phi(G_j)$.

Dynamics: For $t > 0$, the system is updated according to the following dynamics:

1. An iteration of the mesoscale dynamics according to the model described in Section 2.2 is carried out. This iteration produces a population of cells with the corresponding set of different genotypes present in the population at iteration $t = t+1$, $S_G(t+1)$ and with the phenotype set $S_\phi(t+1)$.
2. The growth dynamics of the genotype–phenotype graph is defined as follows:
 - If $S_G(t+1) \cap V_G(t) \neq S_G(t+1)$, i.e. the population dynamics has generated new genotypes, then $V_G(t+1) = V_G(t) \cup (S_G(t+1) - S_G(t+1) \cap V_G(t))$, i.e. we update the set of genotype nodes by adding the new ones arising from the population (mesoscale) dynamics.
 - Similarly, for the set of phenotype nodes, if $S_\phi(t+1) \cap V_\phi(t) \neq S_\phi(t+1)$ then $V_\phi(t+1) = V_\phi(t) \cup (S_\phi(t+1) - S_\phi(t+1) \cap V_\phi(t))$, i.e. the set of phenotype nodes is updated by

Fig. 2. Time evolution of the degree distribution for the genotype nodes in the bipartite network. The degree of the genotype nodes is exponentially distributed. We observe that the genotype degree distribution quickly settles into its exponential distribution steady-state.

addition of the new ones arising from the population dynamics.

- The set of genotype–genotype links, $E_G(t+1)$, is updated by adding to $E_G(t)$ the following set of edges: for all new genotype, $\mathcal{G} \in S_G(t+1) - S_G(t) \cap V_G(t)$, we add new links between $\mathcal{G}$ and all $\mathcal{G}_j \in V_G(t)$ such that $\text{dist}(\mathcal{G}, \mathcal{G}_j) = 1$, i.e. we add links from the new genotypes to all the genotypes in the graph which are one mutation apart.
- The set of genotype–phenotype links, $E_\phi(t+1)$, is updated by applying the genotype–phenotype map (microscale dynamics) to the new genotypes, i.e. new edges are added between phenotype node ϕ_i and genotype node $\mathcal{G}_j$ if and only if $\phi_i = \phi(\mathcal{G}_j)$.

2.4.1. Network dynamics: degree distributions

The result of the above multiscale model is the evolutionary dynamics on the genotype–phenotype space (i.e. our pseudo-bipartite network). This space can be characterised in terms of a

number of topological attributes typical from complex network theory (Newman, 2010). One such metric, which conveys very useful information, is the degree distribution. Here, we consider two different degree distributions: the genotype degree distribution, $P(k_g)$, corresponding to the number of genotype neighbours of genotype nodes, and the phenotype degree distribution, $P(k_\phi)$, corresponding to the number of genotype nodes connected to a phenotype node (i.e. k_ϕ is an approximate measure of the size of the basin of attraction of phenotype ϕ subjected to the selective pressure). Results for $P(k_g)$ and $P(k_\phi)$ are shown in Figs. 2 and 3, respectively.

For GRNs whose topology has been generated according to the Strogatz–Watts model (see Figs. 2 and 3), we observe that, whereas the genotype degree distribution quickly settles into an (steady-state) exponential distribution, the phenotype degree distribution evolves towards a power-law distribution. The same qualitative behaviour is observed for scale-free GRNs, although some significant differences are observed with respect to the previous case. Fig. 4, where we compare the genotype and phenotype degree distributions corresponding to both GRNs

Fig. 3. Time evolution of the phenotype degree distribution in the bipartite network. The degree of the phenotype nodes is distributed according to a power law. We observe that the phenotype degree distribution quickly settles into its power law distribution steady-state.

Fig. 4. These plots show the comparison between the steady-state genotype and the phenotype degree distribution for scale-free GRNs and Strogatz–Watts GRNs. We observe that whereas the corresponding genotype degree distributions, $P(k_g)$, do not differ in a significant way, the phenotype degree distribution, $P(k_\phi)$, for scale-free GRNs exhibits a fatter tail than its Strogatz–Watts counterpart.

topologies, shows that whereas the genotype degree distribution for both scale-free and SW GRNs does not significantly differ, the phenotype degree distribution corresponding to the scale-free GRN exhibits a fatter tail than its Strogatz–Watts counterpart. The significance of this behaviour will become apparent when analysing the robustness properties of evolvability to be discussed in Sections 4 and 5.

Finally, we have considered the evolution of our multi-scale evolutionary model under two distinct regimes, namely, $\mu N > 1$ and $\mu N < 1$. These two regimes corresponding to fundamentally different dynamical regimes (Wagner, 2007, 2008a). Populations with $\mu N > 1$ are very likely to be polymorphic at any given generation, i.e. more than one phenotype coexist within the population, whereas populations with $\mu N < 1$ tend to be monomorphic for most of their evolution. Consequently, whilst the former will evolve to accumulate over time in regions of the genotype–phenotype map where strongly robust genotypes exist, the latter will perform a random walk which will sample the phenotype–genotype space uniformly. Given these differences, it is important to compare the

behaviour of our system in these two regimes. Regarding the genotype degree distribution, $P(k_g)$, and the phenotype degree distribution, $P(k_\phi)$, we observe that they exhibit fatter tails when $\mu N < 1$ than in the case $\mu N > 1$, as illustrated in the simulations shown in Fig. 5. The consequences regarding robustness and evolvability of this fact will be explored in detail in Sections 4 and 5.

3. Topological characterisation of phenotypic robustness

In this section, we provide a definition of a metric of phenotypic robustness based on the topological properties of the genotype–phenotype graph.

Our working definition of phenotypic robustness corresponds to the likelihood of a viable individual to retain its phenotype when gene mutations occur (Wagner, 2007). Our aim in this section is to propose a metric for phenotypic robustness that allows us to quantitatively analyse its evolutionary dynamics Tables 1 and 2.

Fig. 5. Simulation results comparing the steady-state genotype and phenotype degree distributions for different values of the mutation rate, μ . We observe that whereas both the corresponding distributions, $P(k_g)$ and $P(k_\phi)$, exhibit a fatter tail for $\mu N < 1$ than for $\mu N > 1$. We have taken $N=50$ and $\mu = 0.3$ (red dots, $\mu N > 1$) and $\mu = 0.01$ (blue dots, $\mu N < 1$). These results correspond to genotype–phenotype maps generated using Strogatz–Watts GRNs with $p=0.1$. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

Table 1
Parameter values.

Parameter	Description	Typical value
N_G	Length of the genome	20
l_v	Maximum period of viable oscillations	5 or 7
p	Rewiring probability in the Watts–Strogatz model	0.1 or 0.9
E	Number of links in the GRN	40
p_+	Probability of positive feed-back link	0.5
N_c	Number of cells in the population	50
μ	Mutation probability	0.3

Table 2
Parameters.

Parameter	Description
t	Time or iterations
$A = (a_{ij})$	Adjacency matrix of the interaction gene graph without to indicate if an interaction is positive or negative, i.e. $a_{ij} \in \{0, 1\}$
$\mathcal{G} = (G, g(0))$	$G = (g_{ij})$: matrix of interaction between genes ($g_{ij} = \pm 1$), $g(0) = (g_i(0))$: vector of the initial pattern of gene expression ($g_i(0) = \pm 1$)
ϕ	Phenotype and $\phi(\mathcal{G})$: phenotype corresponding to genotype $\mathcal{G}$
B	Pseudo-bipartite graph with: V_G : genotype nodes, V_ϕ : phenotype nodes, E_G : genotype–genotype links, E_ϕ : genotype–phenotype links. $B = (b_{ij})$ adjacency matrix of B
k_g	Genotype degree in the bipartite network
k_ϕ	Phenotype degree in the bipartite network
$P(k_g), P(k_\phi)$	Genotype and phenotype degree distribution in the bipartite network, respectively
c_ϕ	Clustering coefficient of phenotype ϕ
S_{gcc}	Size of the giant component
$\pi_\phi(s)$	Distribution of the size of the connected components
$D_\phi(t)$	Diameter of the phenotype network at time t

According to this definition, the phenotype $\phi(\mathcal{G})$ of an individual with genotype $\mathcal{G}$ is deemed to be robust if a random gene mutation, transforming the genotype $\mathcal{G} \rightarrow \mathcal{G}'$ where $\text{dist}(\mathcal{G}, \mathcal{G}') = 1$,

does not affect the phenotype, i.e. if $\phi(\mathcal{G}) = \phi(\mathcal{G}')$. This definition has a topological equivalence in terms of the properties of the pseudo-bipartite network defined in Section 2.3. Since $\text{dist}(\mathcal{G}, \mathcal{G}') = 1$ these two genotypes correspond to nodes linked by a genotype–genotype edge. Moreover, if $\phi(\mathcal{G}) = \phi(\mathcal{G}')$ this means that both nodes $\mathcal{G}$ and $\mathcal{G}'$ are linked to the same phenotype node ϕ by genotype–phenotype edges. In other words, the genotype nodes $\mathcal{G}$ and $\mathcal{G}'$ and the phenotype node ϕ form a triangle within the pseudo-bipartite graph. Thus the number of such triangles within the neutral network of a given phenotype (i.e. the basin of attraction of ϕ) is a direct measure of phenotypic robustness. Phenotypic robustness can thus be quantified by means of the clustering coefficient of phenotype node ϕ :

$$c_\phi = \frac{2T_\phi}{k_\phi(k_\phi - 1)} \quad (3)$$

where T_ϕ is the number of triangles that have ϕ as one of their vertices and k_ϕ is the degree of ϕ , i.e. the number of genotype nodes to which ϕ is connected.

Note that k_ϕ , i.e. roughly speaking, the size of the basin of attraction of ϕ , is not an appropriate measure of robustness as it does not take into account how tightly interconnected are the genotypes linked to ϕ . In this respect, our characterisation of robustness is reminiscent of the one given by Ciliberti et al. (2007a,b) where robustness is characterised by the community structure within their genotype graph Γ (see Section 2.3). However, we found that, due to known difficulties with the algorithms currently used to determine community structure in graphs (different algorithms produce different communities, miss-assignment of nodes to clusters, etc., Fortunato, 2010), quantifying robustness in terms of community structure is problematic, hence our introduction of this alternative description in terms of the pseudo-bipartite graph.

Fig. 6 shows simulation results regarding the steady-state behaviour of the relation between the clustering coefficient of phenotype nodes, c_ϕ , i.e. phenotypic robustness, as a function of their degree, k_ϕ , the size of the pool of viable genotypes with phenotype ϕ . These results show that there is an inverse

Fig. 6. Blue dots represent simulation results for the (steady-state) phenotype clustering coefficient as a function of the degree, k_ϕ . We compare with $c(k_\phi) = (k_\phi - 1)^{-1}$ (solid black line), which according to Serrano and Boguñá (2006a,b) establish the border between high and low clustering regimes in uncorrelated networks. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

correlation between c_ϕ and k_ϕ : low-degree phenotypes exhibit much higher clustering coefficient than high-degree phenotypes.

We further observe that by tuning the rewiring probability of the GRN, p , the overall robustness can be controlled. Fig. 6 shows that for $p=0.9$, regardless of the value of the of l_v , $c_\phi(k_\phi)$ is very well close to $c_\phi(k_\phi) \simeq (k_\phi - 1)^{-1}$, whereas for $p=0.1$, the corresponding $c_\phi(k_\phi)$ curve is such that $c_\phi(k_\phi) > (k_\phi - 1)^{-1}$. This implies that the robustness of the phenotypes generated with GRNs with low values of p is intrinsically more robust. The reason for this result is that the parameter p in SW model measures how far the topology of the network is from a regular lattice and how close is to a tree. Low values of p imply that the network closely resembles a regular lattice. This means that low- p networks are, generally speaking, more heavily clustered than networks with large p values, which are more tree-like. More clustered GRNs imply that the presence of feed-back loops is more likely which, in turn, implies that oscillatory phenotypes are more common. Therefore, it is less probable to obtain viable phenotypes for SW-GRNs with large p . However, when one is discovered, it is very likely that it is more robust.

4. Topological characterisation of evolvability

Evolvability is defined as the ability of an evolutionary system to innovate by generating new and better adapted behaviour (Wagner, 2007). Recent results have shed some light into the apparent conflict between robustness, i.e. resilience to change, and evolvability (Ciliberti et al., 2007a,b; Wagner, 2008a,b, 2011, 2012). These results suggest that, under very general conditions, the tension is only apparent and the mechanisms that favour robustness also enforce the emergence of evolvability.

Regarding innovation, we follow the work of Ciliberti et al. (2007a,b) where evolvability is related to the ability of the system to, under the effect of gene mutations, move from one neutral network (i.e. the subnetwork of genotypes with one particular phenotype) to another. This view, that evolvability is equivalent to the navigability of the network of phenotypes, has a direct mathematical translation within our model.

In this section, we first define the phenotype network as the (weighted) one-component projection of the pseudo-bipartite genotype–phenotype graph. We then go on to characterise the

emergence of a giant connected component within the phenotype network, an attribute that we use to characterise evolvability in topological terms. Finally, we conclude this section by showing that the phenotype network exhibits the small-world network property.

4.1. Phenotype network

In this and the following sections, we consider the properties of the phenotype network, which is defined as a one-mode projection of our phenotype–genotype network, which is obtained in the following way. We consider that between a pair of phenotypes, ϕ_i and ϕ_j , there is an edge if there exists a path $\phi_i - \mathcal{G}_i - \mathcal{G}_j - \phi_j$ where $\phi(\mathcal{G}_i) = \phi_i$ and $\phi(\mathcal{G}_j) = \phi_j$, and $\mathcal{G}_i$ and $\mathcal{G}_j$ are connected within the pseudo-bipartite network. In other words, two phenotype nodes are connected in the one-component projection if there exists at least one length 3 path between them in pseudo-bipartite network. In our model such length 3 paths represent the minimum (two mutation) evolutionary path to reach a phenotype from another. Since there any number of different length 3 paths between any pair of phenotypes, we define the phenotype network as a weighted network where the weight of the edge

between two phenotypes is the number of such paths that connect them within the full genotype–phenotype network.

Mathematically, the phenotype network is constructed as follows. If $B = (b_{ij})$ is the adjacency matrix of the pseudo-bipartite graph, then in order to calculate the number of length 3 paths between phenotypes, we use a well-known property of the adjacency matrix, namely that the entries of the matrix $B^3 = (\beta_{ij})$ are the number of length 3 paths between nodes i and j . β_{ij} is zero if no path of length 3 exists between nodes i and j . By considering only the entries corresponding to paths between phenotype nodes, we obtain the weighted adjacency matrix corresponding to our phenotype network.

4.2. Evolution of evolvability: emergence of giant connected component

The onset of percolation is heralded by the formation of the so-called giant connected component. In percolation theory (Chirstensen and Moloney, 2005), the giant component is a connected subset of nodes that contains a macroscopic fraction of the entire set of vertices of the graph, i.e. $s_{gcc} = \gamma N_v$, where s_{gcc} is

Fig. 7. Time evolution of the distribution of cluster size in the phenotype network.

Fig. 8. Dynamics of the formation of giant connected component as shown by the emergence of a peak in the distribution of cluster size in the phenotype network. The shadowed region is the $1-\sigma$ confidence region using a Gaussian kernel density estimation.

the size of the giant component, N_v is the number of vertices of the graph and $\gamma \lesssim 1$.

In order to address the question of whether the evolutionary dynamics of our multiscale system gives rise to evolvability, i.e. to a phenotype (network) space where a macroscopically large number of all viable phenotypes are mutually connected, we study the time evolution of the distribution of the size of the connected components, $\pi_\phi(s)$, in the phenotype network. The results are shown in Fig. 7.

We observe that, regardless the parameter values, as time progresses the formation of a giant connected component emerges. The formation of this connected component is signalled by the appearance of a peak for large size clusters in the cluster size distribution. Whereas the emergence of the giant connected component is unaffected by the GRN parameter, p , or the strength of the selective pressure, l_v , its dynamic is indeed affected by the rewiring probability: the larger the value of p , the faster the formation of the giant connected component (see Fig. 8).

4.3. The phenotype network is a small world

Complex networks often exhibit the so-called small-world phenomenon (Watts and Strogatz, 1998). Small-world networks are such that most nodes can be reached from every other node in a small number of jumps compared to the size of the network. Mathematically, this property is defined by requiring that the diameter of the network grows proportional to the logarithm of the number of nodes (Newman, 2010). We are interested in investigating whether the evolutionary process described in Section 2 yields an evolving phenotype network with the small-world property, i.e. whether

$$D_\phi(t) \sim \log N_\phi(t) \quad (4)$$

where $D_\phi(t)$ is the diameter of the phenotype network at time t and $N_\phi(t)$ is the number of nodes at time t .

Simulation results regarding the time evolution of both the diameter of the phenotype network and the number of nodes (i.e. the number of viable phenotypes that our evolutionary dynamics

Fig. 9. Diameter of the growing phenotype network as a function of the logarithm of the number of nodes. The linear dependence of the diameter of the phenotype network on the logarithm of the number of nodes shows that the phenotype network exhibits small-world behaviour.

Fig. 10. Damage suffered by the phenotype giant connected component. The relative size of the giant connected, %gcc, is defined as the cardinal of the set of nodes such that, for any pair of nodes, it is possible to find at least a continuous connecting path relative to the total number of nodes. Damage is measured in terms of the relative size of the remaining largest connected component after removal of a certain percentage of edges. We consider that edges are (a) removed randomly, (b) according to the clustering coefficient of the nodes (phenotypes) to which they are connected, and (c) according to the degree of the phenotype to which they are connected. Clustering coefficient and degree of the phenotypes refer to the quantities corresponding to the bipartite genotype–phenotype network. Plots (d), (e) and (f) explore the comparison between the behaviour of the phenotype giant connected component generated by a multiscale model with Strogatz–Watts (SW) and scale-free (SF) gene regulatory networks.

(see Section 2) allows to emerge) are shown in Fig. 9. We observe that the networks of phenotypes generated by our evolutionary model have the small world property as their diameter grows linearly with the logarithm of the number of phenotypes (nodes). This result is stronger than the similar one obtained recently by Aguirre et al. (2011) regarding the genotype–phenotype map for RNA secondary structures. They have been demonstrated that the average path length within RNA secondary structure neutral networks scales logarithmically with the size of each independent neutral network. Here, we are able to show that our multi-scale model generates a small-world phenotype network.

From these simulation, we observe that our results regarding the small-network structure of the phenotype network is robust to changes in several parameter values. For example, according to Fig. 9, modifying both p , i.e. the rewiring probability of the SW model used to generate the GRN topology, and l_v , essentially, the selective pressure, leaves untouched the small-world property.

Taken together, these two properties, namely, the emergence of a giant connected component within the phenotype network and of the small-world phenomenon, have strong biological implications, in particular regarding evolvability. First, the fact that a giant connected component in phenotype space exists means that it is globally connected and, therefore, innovation by means of genetic mutations is granted regardless of the existence of robust phenotypes. This is consistent with previous results regarding innovation and robustness (Wagner, 2007, 2008a,b, 2012; Ciliberti et al., 2007a,b). On the other hand, the small world property of our phenotype network implies that, in topological distance, phenotypes are very close to each other, and therefore, adaptive processes where the system must reach a privileged phenotype in order to ensure survival (as, for example, in evolutionary escape Iwasa et al., 2003, 2004) could be much more efficient than previously thought.

5. Robustness of evolvability

We have shown in Section 4 that similar to what has been observed in other models (Wagner, 2007, 2008a,b, 2012; Ciliberti et al., 2007a,b) evolvability is an emergent, evolved property of the dynamics. This view in which evolvability is itself an evolutionary property of the dynamics (Pigliucci, 2008), together with our topological characterisation of robustness, allows us to pose the question of whether evolvability itself is a robust property.

Fig. 12. Average nearest neighbours as a function of the phenotype degree, k_ϕ , for different values of the Strogatz–Watts GRN parameters, l_v and p . Including a zoom of the figure.

Fig. 11. Variance of the giant component size from damage done when edges are removed randomly, according to the clustering coefficient of the nodes (phenotypes) to which they are connected, and according to the degree of the phenotype to which they are connected. Clustering coefficient and degree of the phenotypes refer to the quantities corresponding to the bipartite genotype–phenotype network.

In order to address the issue of robustness of evolvability, we proceed in the following way. Ciliberti et al. (2007a,b) have shown that evolvability is associated with the connectivity between neutral networks corresponding to different phenotypes. In Section 4.2, we have shown that this interpretation of evolvability leads to its characterisation in terms of the emergence of a giant connected component in the growing phenotype network. Within this framework, a natural definition of robustness of evolvability arises: the robustness of evolvability is associated with the resilience against edge-removal of the corresponding giant connected component.

In order to ascertain how different factors affect the robustness of evolvability, i.e. the integrity of the giant connected component, we consider three different strategies for edge-removal (Callaway et al., 2001): random removal and two targeted strategies, one in which removal of edges is done according to the clustering phenotypes to which they are connected, and the other in which removal of edges according to the degree of the phenotypes to which they are connected. Note that here clustering coefficient and degree of the phenotypes refer to the corresponding quantities in the bipartite genotype–phenotype network. Results are shown in Fig. 10.

Regarding the behaviour of the phenotype giant component generated by Strogatz–Watts GRNs, Fig. 10(a) shows that random edge removal has little effect on the integrity of the giant connected component unless a massive percentage of edges is removed. This implies that evolvability is very robust to random elimination of genotypes. By contrast, when edges are removed in targeted manner according to the degree of the phenotype, we observe that there is a sharp transition when the percentage of edges removed exceeds a critical point (see Fig. 10(c)). This means that evolvability is not robust to targeted attack where genes associated with high-degree phenotypes are the focus of the attack.

The behaviour described so far is very much resembling the behaviour of other complex networks, namely, robust against random attack but sensitive to high-degree-targeted attack (Callaway et al., 2001). We consider a third scenario in which the attack is targeted on the more robust phenotypes, i.e. those phenotypes with larger clustering coefficient, c_ϕ (Eq. (3)). In this case, the behaviour of the system is less clearcut.

Fig. 11 shows the variance of the size of phenotype giant connected component, $\sigma_{g_{cc}}^2$, for Strogatz–Watts GRNs. We see that, for the two targeted strategies, this quantity exhibits the sharp increase typical of a second-order phase transition. It is worth noting that the threshold/critical value for the giant connected component to disappear is not significantly different regardless of which targeting strategy is used.

Regarding the effects of the Strogatz–Watts GRN parameters, our simulation results show that evolvability is more robust in systems with larger values of the rewiring probability p . In this respect, it is worth noticing that, in opposite to the behaviour observed in clustered uncorrelated networks, analysed in depth in Serrano and Boguñá (2006a,b), we observe that for $p=0.1$, for which $c_\phi(k_\phi) \leq (k_\phi - 1)^{-1}$ (as shown in Fig. 6), the giant connected component is less robust than for $p=0.9$, for which $c_\phi(k_\phi) \simeq (k_\phi - 1)^{-1}$ (see in Fig. 6). This difference in behaviour is due to the presence of correlations.

The presence of correlations is demonstrated by analysing the average number of nearest-neighbours as a function of the degree, $k_{nn}(k_\phi)$, which is shown in Fig. 12. These results show that $k_{nn}(k_\phi)$ is an increasing function of the phenotype degree and, therefore, the phenotype network is assortative (more connected nodes are preferably connected to each other). Note that, although this is in contrast to the common situation in biological networks lacking an underlying genotype–phenotype structure, which are commonly

found to be dissortative (more connected nodes are preferably connected to low-degree nodes) (Newman, 2010), assortativity seems to be a systematic property of phenotype networks. A recent example is provided in Cowperthwaite et al. (2008) where it is observed that large (i.e. strongly connected) phenotypes have an inherently enhanced accessibility to new phenotypes.

It is also interesting to compare the behaviour of the robustness of the phenotype giant connected component generated by Strogatz–Watts (SW) GRNs and by scale-free (SF) GRNs. Fig. 10 (d), (e) and (f) show that the giant connected component corresponding to SF GRNs is systematically more robust than those generated by SW GRNs.

We now address the effect of varying the mutation rate on our results regarding robustness of evolvability. We have argued in Section 2.4.1 that there are two distinct regimes, corresponding to $\mu N > 1$ and $\mu N < 1$ where the population exhibits rather different structure: whilst for $\mu N > 1$ the population is most likely to be polymorphic, for $\mu N < 1$ the population is monomorphic most of the time. We observe that, in agreement with results reported by Wagner (2008a) regarding his analysis of the RNA genotype–phenotype map, where evolvability is shown to be robust to variations in the mutation rate, our multi-scale model exhibits a similarly robustly connected phenotype network. In fact, the phenotype giant connected component, i.e. the core of mutually reachable phenotypes, is found to be more resilient to damage for small mutation rates (in the $\mu N < 1$ regime), regardless of the targeting strategy we use. Increased robustness of the phenotype giant connected component for smaller mutation rate is a consequence of the presence of fatter tails in the genotype and phenotype degree distributions as shown in Section 2.4.1, which implies that the giant connected component corresponding to $\mu N < 1$ is more tightly connected and, therefore, more difficult to disconnect (Fig. 13).

6. Conclusions

In this paper we have presented a novel mathematical description of the genotype–phenotype space in terms of a pseudo-bipartite graph and its associated one-component projection, the phenotype network. We have based our presentation on a model of the genotype–phenotype for circuits of gene regulation, although its extension to other genotype–phenotype models, such as RNA, should be straightforward. This new representation allows us to characterise robustness of evolvability in terms of the topological properties of these networks: phenotypic robustness is defined as the clustering coefficient of phenotype nodes in the pseudo-bipartite genotype–phenotype network, and evolvability is defined as the emergence of a giant connected component in the phenotype network, which ensures global connectedness and navigability of the space of phenotypes.

New results have been obtained regarding the small-world property of the phenotype network, as well as the characterisation of the robustness properties of evolvability. In particular, we have shown that the phenotype network exhibits, beyond global connectedness, the small-world property. We have further exploited our topological definition of evolvability, characterised in terms of the onset of percolation and the size distribution of connected components, to explore the issue of whether evolvability is a robust property, i.e. under which conditions the giant connected component may be broken down. We observe, that under random attack, the giant connected component, and, therefore, our system's evolvability, is very resilient. However, similar to other complex networks, when the attack is targeted rather than random, weaknesses arise and evolvability breaks down.

Our results regarding the topological characterisation of robustness and evolvability are complementary to the corresponding definitions given by Ciliberti et al. (2007a,b). By extending their

description in terms of a space of genotypes to explicitly include the corresponding phenotypes, we can provide purely topological definitions in terms of easily measurable quantities for which algorithms are readily available (Newman, 2010), namely, the clustering coefficient of the phenotype nodes as a measure of phenotypic robustness, and the existence of a giant connected component in the phenotype network as characteristic of evolvability.

Regarding the characterisation of robustness, our results given in Section 3, Fig. 6, regarding the dependence of the phenotype clustering coefficient on the degree of the phenotype nodes, $c_\phi(k_\phi)$, show that there exists an inverse relation between clustering coefficient and degree: $c_\phi(k_\phi) \propto (k_\phi - 1)^{-\alpha_{p,l_v}}$ with $\alpha_{p,l_v} \geq 1$. This result has an important interpretation in terms of the concept of *neighbourhood* and the relation between robustness and evolvability (Wagner, 2011, 2012). Our result implies that those phenotypes with larger cryptic variability, i.e. genotypic variability which presents no variation in phenotype and which is given by the degree of the phenotype nodes in the pseudo-bipartite genotype-phenotype network, k_ϕ , exhibit, in relative terms, higher accessibility to new phenotypes than those with smaller degree. This observation suggests that these phenotype nodes have a fundamental role to play in the ability for innovation of the system.

Our analysis of the robustness of evolvability is closely related to this issue. We have shown (see Section 5, Fig. 10) that evolvability is a robust property with respect to attack (removal) of randomly selected genotypes. Recall that removal of a genotype node in the genotype-phenotype network corresponds to the removal of an edge in the phenotype network. Mathematically, this corresponds to the giant connected component in the phenotype network being robust against random removal of edges, Fig. 10(a). This situation is radically changed when genotype removal is targeted rather than random. In particular, we have carried out two removal strategies, namely, remove first those genotypes connected to phenotypes with bigger clustering coefficient (Fig. 10(b)), and remove first those genotypes connected to phenotypes with bigger degree (Fig. 10(c)). In other words, we target genotypes associated with more robust phenotypes or more connected phenotypes. These targeted attacks are much more efficient at breaking the giant connected component. In particular, evolvability is particularly affected by the removal of genotypes associated with phenotypes with bigger degree, k_ϕ . Removal of genotypes associated with phenotypes with bigger clustering coefficient, c_ϕ , seems to be slightly less detrimental, although its effect on evolvability appears to be much more important than that associated with random attack.

Taken together, these results support the theory that robustness and cryptic variability, rather than hindering innovation, they facilitate the ability of evolutionary systems to evolve and adapt. These results are, therefore, in agreement with recent findings by Wagner (2011, 2012).

Evolvability is further quantified by the distribution of size of connected components (see Section 5). This measure contains a much more detailed description of evolvability: it provides the probability of a phenotype to be connected to a cluster (i.e. a set of mutually reachable phenotypes) of size s . This distribution provides a description of evolvability as it gives us information about

Fig. 13. Damage suffered by the phenotype giant connected component when edges are (a) removed randomly, (b) according to the clustering coefficient of the nodes (phenotypes) to which they are connected, and (c) according to the degree of the phenotype to which they are connected. Clustering coefficient and degree of the phenotypes refer to the quantities corresponding to the bipartite genotype-phenotype network. Green lines correspond to $\mu = 0.01$ (i.e. $\mu N = 0.5$) whereas red lines correspond to $\mu = 0.3$ (i.e. $\mu N = 1.5$). These phenotype networks are generated by a multiscale model with Strogatz-Watts with $p=0.1$. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

how many new phenotypes are in reach of a given phenotype. It also provides a measure of evolvability before the giant connected component has been formed (as shown in Fig. 7) and, equivalently, after the damage to the phenotype network has been done. If we look at Fig. 7, for example, we realise that before the formation of the giant connected component some degree of evolvability is still possible. Taken together both topological measures, existence of a giant connected components and distribution of size of connected components provide a characterisation of evolvability which allows us to explore whether evolvability is at all possible and whether a macroscopic proportion of phenotypes are mutually reachable.

A new result that emerges from our analysis is the fact that the genotype–phenotype map is such that the phenotype network exhibits the small-world phenomenon (see Section 4, Fig. 9). This result implies that the maximum average distance between two given phenotypes on the phenotype network scales as the logarithm of the number of phenotypes, which is much smaller than the distance one should expect if the phenotype–genotype space were lattice-like, as it is commonly assumed in models of evolutionary escape, for example Iwasa et al. (2003, 2004). This topological result has obvious implications regarding the rate of evolutionary adaptation, which should be greatly increased by this property. This result is consistent too with recent studies suggesting that cryptic genetic variation increases the rate of evolutionary adaptation in RNA enzymes (Hayden et al., 2011).

Besides the issues arising from our topological characterisation of robustness and evolvability and their biological relevance, we have also investigated their controllability, i.e. to what extent change of the parameters of the GRNs can be used to drive changes in robustness and evolvability. Although our results show that the generic properties of the genotype–phenotype space are rather insensitive to these parameters, we have found that some level of control is indeed possible. In particular, in the case of GRN topology generated using the Strogatz–Watts model, from our results shown in Fig. 6, we can conclude that a measure of control on the level of robustness can be achieved by tuning the characteristic parameters of the GRN. Increasing the rewiring probability p appears to lower the overall phenotypic robustness, as shown in Fig. 6 where we observe that the clustering coefficient as a function of phenotype degree for $p=0.1$ is increased with respect to the corresponding values for $p=0.9$. We have also investigated how some model details affect our results. In particular, we have explored the effect of (i) considering a scale-free GRN topology, using the Barabási–Albert model instead of the Strogatz–Watts model, and (ii) varying the value of the mutation rate, μ . We have observed both considering scale-free GRN topology and decreasing μ render the evolvability more robust, as both these factors produce phenotype networks with giant connected components that are more resilient to damage.

A biological scenario where our topological characterisation of robustness and evolvability could be a useful tool is cancer. In a recent review, Tian et al. (2011) have advocated that in order to understand the principles of tumour evolvability, which, in turn, are key to analyse issues such as resistance to therapy and cancer stem cells, a systems biology approach must be used which integrates the different factors that contribute to heritable phenotypic variability (genetic and epigenetic instability, stochastic protein dynamics, tumour microenvironment, etc.). Central to these issues is the concept of *epigenetic landscape*, which are inspired by the concept of energy landscape familiar in Physics: a space populated with diverse attractor states (i.e. minima of the landscape), corresponding to different cell states and separated by *epigenetic barriers*. In the context of our model, the epigenetic landscape is represented by the phenotype network.

In developmental biology, these attractors correspond to the stable-steady states of the GRN that regulate cell differentiation

and represent the differentiation states of the cell (Huang, 2009; Huang et al., 2009; Huang, 2011; Villarreal et al., 2012), and transitions between these states occur following a well-orchestrated series of events, otherwise cells are locked into the corresponding state. The epigenetic landscape is organised so that (pluripotent) stem cells sit atop a hierarchy of connected attractors that radiate outwards to stable states which represent distinct cell fates.

In the case of cancer cells, pathological attractors that represent different cancer states are scattered over the epigenetic landscape. In some cases, cancer cells possess a landscape with well-defined minima and correspondingly stable cancer states similar to developmental landscape. However, due to severe genetic and epigenetic dysregulations occurring in aggressive forms of cancer, it appears that a de-stabilisation of the epigenetic landscape ensues where epigenetic barriers are lowered down and attractors becoming increasingly less stable, thus increasing evolvability and the ability of the tumour to evade therapy. Supporting evidence for this model (reviewed in Tian et al., 2011) has been recently found (Roesch et al., 2010; Sharma et al., 2010) where genetic and epigenetic dysfunctions allow for the population of cancer cells to transiently visit a number of metastable states within the epigenetic landscape, thus disrupting the hierarchical organisation of normal tissues and favouring the emergence of phenomena such as transient drug resistance (Sharma et al., 2010). The application of the formalism developed to such issues would require to develop a model of the GRN governing such states as well as the corresponding epigenetic regulation (Roesch et al., 2010) and then analyse using our topological description (based on percolation on and distribution of component size in the phenotype network) how dysregulations at the genetic and epigenetic levels would affect evolvability, whereby targeted strategies aimed at phenotypes which cause the most disruption on evolvability and produce more static tumours that could be formulated. The formulation of such models and the corroboration of whether such strategies are based on topological properties (like the ones proposed in Section 5) or not is beyond the scope of this work and left for future research.

Finally, the results presented in this work open several avenues for future work. For example, the issues arising from the small-world property of the phenotype network, in particular those regarding its effects on adaptability and evolutionary escape need a more in-depth examination. Another aspect in which further research is to be conducted is the increased resilience of evolvability under increased selective pressure (corresponding to larger values of l_v) and increased randomisation of the GRN topology (corresponding to larger values of the rewiring probability, p), as observed in Fig. 10.

Acknowledgements

The authors gratefully acknowledge the Spanish Ministry for Science and Innovation (MICINN) for funding under Grant MTM2011-29342 and Generalitat de Catalunya for funding under Grant 2009SGR345. The authors are also grateful to two anonymous referees whose comments greatly improved this manuscript.

References

- Aguirre, J., Buldú, J.M., Stich, M., Manrubia, S.C., 2011. Topological structure of the space of phenotypes: the case of RNA neutral networks. *PLoS One* 6, e26324.
- Alarcón, T., Byrne, H.M., Maini, P.K., 2005. A multiple scale model of tumour growth. *Multiscale Model. Simul.* 3, 440–475.
- Albert, R., 2005. Scale-free networks in cell biology. *J. Cell Sci.* 118, 4947–4957.

- Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M., Teichmann, S.A., 2004. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.* 14, 283–291.
- Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286, 509–512.
- Bergman, A., Siegal, M.L., 2003. Evolutionary capacitance is a general feature of complex gene networks. *Nature* 424, 549–552.
- Blythe, R.A., McKane, A.J., 2007. Stochastic models of evolution in genetics, ecology and linguistics. *J. Stat. Mech.*, P07018.
- Callaway, D.S., Hopcroft, J.E., Kleinberg, J.M., Newman, M.E.J., Strogatz, S.H., 2001. Are randomly grown graphs really random? *Phys. Rev. E* 64, 041902.
- Christensen, K., Moloney, N.R., 2005. Complexity and Criticality. Oxford University Press, New York, USA.
- Ciliberti, S., Martin, O.C., Wagner, A., 2007a. Robustness can evolve gradually in complex regulatory gene networks with varying topology. *PLoS Comput. Biol.* 3, e15.
- Ciliberti, S., Martin, O.C., Wagner, A., 2007b. Innovation and robustness in complex regulatory gene networks. *Proc. Natl. Acad. Sci.* 104, 13591–13596.
- Cowperthwaite, M.C., Economou, E.P., Harcombe, W.R., Miller, E.L., Meyers, L.A., 2008. The ascent of the abundant: how mutational networks constrain evolution. *PLoS Comput. Biol.* 4, e1000110.
- Eyre-Walker, A., Keightley, P.D., 1990. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 61, 610–618.
- Fontana, W., Schuster, P., 1998. Shaping space: the possible and the attainable in RNA genotype-phenotype mapping. *J. Theor. Biol.* 194, 491–515.
- Fortunato, S., 2010. Community detection in graphs. *Phys. Reports* 486, 75–174.
- Sam F. Greenbury, Iain G. Johnston, Matthew A. Smith, Jonathan P.K. Doye, Ard A. Louis, 2010. The effect of scale-free topology on the robustness and evolvability of genetic regulatory networks. *J. Theor. Biol.* 267, 48–61.
- Hayden, E.J., Ferrara, E., Wagner, A., 2011. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474, 9295.
- Huang, S., 2009. Reprogramming cell fates: reconciling rarity and robustness. *BioEssays* 31, 546–560.
- Huang, S., 2011. The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology. *BioEssays* 34, 149–157.
- Huang, S., Ernberg, I., Kauffman, S., 2009. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* 20, 869–876.
- Iwasa, Y., Michor, F., Nowak, M.A., 2003. Evolutionary dynamics of escape from biomedical intervention. *Proc. R. Soc. B* 270, 2573–2578.
- Iwasa, Y., Michor, F., Nowak, M.A., 2004. Evolutionary dynamics of invasion and escape. *J. Theor. Biol.* 226, 205–214.
- Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Assad-Garcia, N., Glass, J.L., Covert, M.W., 2012. A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389–401.
- Kauffman, S.A., 1993. The Origins of Order. Oxford University Press, New York, USA.
- Kitano, H., 2004. Cancer as a robust system: Implications for cancer therapy. *Nat. Rev. Cancer* 4, 227–235.
- Lipman, D., Wilbur, W., 1991. Modelling neutral and selective evolution of protein folding. *Proc. R. Soc. Lond. B* 245, 7–11.
- Ndifon, W., Plotkin, J.B., Dushoff, J., 2009. On the accessibility of adaptive phenotypes of a bacterial metabolic network. *PLoS Comput. Biol.* 5, e1000472.
- Newman, M.E.J., 2010. Networks: An Introduction. Oxford University Press, Oxford, UK.
- Newman, M.E.J., Watts, D.J., 1999. Renormalisation group analysis of the small-world network model. *Phys. Lett. A* 260, 341–346.
- Pigliucci, M., 2008. Is evolvability evolvable? *Nat. Rev. Genet.* 9, 75–82.
- Rodrigues, J.F., Wagner, A., 2009. Evolutionary plasticity and innovations in complex metabolic reaction networks. *PLoS Comput. Biol.* 5, e1000613.
- Rodrigues, J.F., Wagner, A., 2011. Genotype networks in sulfur metabolism. *BMC Syst. Biol.* 5, 39.
- Roesch, A., Fukunaga-Kalabis, M., Schmidt, E.C., Zabierowski, S.E., Brafford, P.A., Vultur, A., Basu, D., Gimotty, P., Vogt, T., Herlyn, M., 2010. A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell* 141, 583–594.
- Samal, A., Rodrigues, J.F., Jost, J., Martin, O.C., Wagner, A., 2010. Genotype networks in metabolic reaction spaces. *BMC Syst. Biol.* 4, 30.
- Sawyer, S.A., Parsch, J., Zhang, Z., Hartl, D.L., 2007. Prevalence of positive selection among nearly neutral amino acid replacement in *Drosophila*. *Proc. Natl. Acad. Sci.* 104, 6504–6510.
- Schuster, P., Fontana, W., Stadler, P., Hofacker, I., 1994. From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond. B* 255, 279–284.
- Serrano, M.Á., Boguñá, M., 2006a. Clustering in complex networks I: general formalism. *Phys. Rev. E* 74, 056114.
- Serrano, M.Á., Boguñá, M., 2006b. Clustering in complex networks II: percolation properties. *Phys. Rev. E* 74, 056115.
- Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., Wong, K.-K., Brandstetter, K., Witner, B., Ramaswamy, S., Classon, M., Settleman, J., 2010. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* 141, 69–80.
- Siegal, M.L., Bergman, A., 2002. Waddington's canalization revisited: developmental stability and evolution. *Proc. Natl. Acad. Sci.* 99, 10528–10532.
- Soskine, M., Tawfik, D.S., 2010. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* 11, 572–582.
- Tian, T., Olson, S., Whitacre, J.M., Harding, A., 2011. The origin of cancer robustness and evolvability. *Integr. Biol.* 3, 17–30.
- Villareal, C., Padilla-Longoria, P., Alvarez-Buylla, E., 2012. General theory of genotype to phenotype mapping: derivation of epigenetic landscapes from N-node complex gene regulatory networks. *Phys. Rev. Lett.* 109, 118102.
- von Noort, V., Snel, B., Huynen, M.A., 2004. The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. *EMBO Reports* 5, 280–284.
- Waddington, C.H., 1942a. Canalisation of development and the inheritance of acquired characters. *Nature* 150, 563–565.
- Waddington, C.H., 1942b. The epigenotype. *Endeavour* 1, 18–20.
- Wagner, A., 1996. Does evolutionary plasticity evolve? *Evolution* 50, 1008–1023.
- Wagner, A., 2007. Robustness and Evolvability in Living Systems. Princeton University Press, Princeton, NJ, USA.
- Wagner, A., 2008a. Neutralism and selectionism: a network-based reconciliation. *Nat. Rev. Genet.* 9, 965–974.
- Wagner, A., 2008b. Robustness and evolvability: a paradox resolved. *Proc. R. Soc. B* 275, 91–100.
- Wagner, A., 2011. Genotype networks shed light on evolutionary constraints. *Trends Ecol. Evol.* 26, 577–583.
- Wagner, A., 2012. The role of robustness in phenotypic adaptation and innovation. *Proc. R. Soc. B* 279, 1249–1258.
- Watts, D.J., Strogatz, S., 1998. Collective dynamics of “small-world” networks. *Nature* 393, 440–442.
- Yook, S.-H., Oltvai, Z.N., Barabási, A.L., 2004. Functional and topological characterization of protein interaction networks. *Proteomics* 4, 928–942.