

Evolutionary dynamics on multiple scales: a quantitative analysis of the interplay between genotype, phenotype, and fitness in linear genetic programming

Ting Hu · Joshua L. Payne · Wolfgang Banzhaf ·
Jason H. Moore

Received: 15 September 2011 / Revised: 12 March 2012 / Published online: 3 May 2012
© Springer Science+Business Media, LLC 2012

Abstract Redundancy is a ubiquitous feature of genetic programming (GP), with many-to-one mappings commonly observed between genotype and phenotype, and between phenotype and fitness. If a representation is redundant, then neutral mutations are possible. A mutation is phenotypically-neutral if its application to a genotype does not lead to a change in phenotype. A mutation is fitness-neutral if its application to a genotype does not lead to a change in fitness. Whether such neutrality has any benefit for GP remains a contentious topic, with reported experimental results supporting both sides of the debate. Most existing studies use performance statistics, such as success rate or search efficiency, to investigate the utility of neutrality in GP. Here, we take a different tack and use a measure of robustness to quantify the neutrality associated with each genotype, phenotype, and fitness value. We argue that understanding the influence of neutrality on GP requires an understanding of the distributions of robustness at these three levels, and of the interplay between robustness, evolvability, and accessibility amongst genotypes, phenotypes, and fitness values. As a concrete example, we consider a simple linear genetic programming system that is amenable to exhaustive enumeration and allows for the full characterization of these quantities, which we then relate to the

Ting Hu and Joshua L. Payne contributed equally to this work.

T. Hu · J. L. Payne (✉) · J. H. Moore
Computational Genetics Laboratory, Dartmouth Medical School, Hanover, NH, USA
e-mail: Joshua.L.Payne@Dartmouth.edu

T. Hu
e-mail: Ting.Hu@Dartmouth.edu

J. H. Moore
e-mail: Jason.H.Moore@Dartmouth.edu

W. Banzhaf
Department of Computer Science, Memorial University of Newfoundland, St. John's, NL, Canada
e-mail: banzhaf@mun.ca

dynamical properties of simple mutation-based evolutionary processes. Our results demonstrate that it is not only the distribution of robustness amongst phenotypes that affects evolutionary search, but also (1) the distributions of robustness at the genotypic and fitness levels and (2) the mutational biases that exist amongst genotypes, phenotypes, and fitness values. Of crucial importance is the relationship between the robustness of a genotype and its mutational bias toward other phenotypes.

Keywords Accessibility · Coreness · Evolvability · Genotype-phenotype map · Phenotype-fitness map · Networks · Neutrality · Redundancy · Robustness

1 Introduction

Redundant mappings between genotype and phenotype are common in genetic programming (GP), where many mutational variants of a genotype yield identical phenotypes [40]. Redundant mappings between phenotype and fitness are also common, with multiple phenotypes producing identical fitness values [43]. Redundant mappings allow for neutrality [1]. In the genotype-phenotype map, a mutation is neutral if its application to a genotype does not lead to a change in phenotype (referred to as phenotypically-neutral). In the phenotype-fitness map, a genetic mutation is neutral if it does not affect fitness (referred to as fitness-neutral). Redundancy and neutrality are thus separate, but related concepts. While neutrality requires redundancy, redundancy does not guarantee neutrality.

Based on a parsimonious model of evolutionary dynamics [40], it has been argued that the potential benefits of redundancy hinge on two distinctions. The first is whether the genotype-phenotype map is uniformly or non-uniformly redundant. A mapping is uniformly redundant if each phenotype is represented by the same number of genotypes and non-uniformly redundant otherwise. The second distinction is whether the genotype-phenotype map is synonymously or non-synonymously redundant. A mapping is synonymously redundant if the genotypes that map to the same phenotype are similar to one another and non-synonymously redundant otherwise. The results of [40] suggest that a non-uniformly redundant mapping is only advantageous in an evolutionary search if the optimal phenotype is overrepresented, while a uniformly redundant mapping offers no advantage. Further, non-synonymously redundant mappings can frustrate evolutionary search because they do not allow recombination operators to work properly.

The neutrality of a redundant mapping, if it exists, can be characterized as a genotype network (a.k.a. a neutral network). In such networks, vertices represent genotypes and edges connect genotypes that share the same phenotype and can be interconverted via single mutational events¹ [51]. By partitioning genotype space into distinct genotype networks, it is possible to provide a more detailed description of redundant representations, complementing the information provided by the

¹ The structure of a genotype network depends upon the type of mutation under consideration. Here, as in related studies [6, 8, 51], we are concerned with point mutations.

uniformity- and synonymity-based classification scheme. For example, within a non-uniformly redundant mapping, the overrepresented phenotype may comprise a single genotype network or a set of several independent genotype networks. Within a non-synonymously redundant mapping, the disparate genotypes of a given phenotype may be connected via a series of phenotypically-neutral point mutations or they may be completely isolated from one another. Clearly these different scenarios have implications for evolutionary search.

In biological systems, genotype networks are often used to describe the neutrality of redundant mappings in terms of robustness [50]. One of the many definitions of robustness is resilience to genetic change, which can be measured using genotype networks. Specifically, the robustness of a genotype is linearly proportional to the number of connections it possesses in the genotype network. The robustness of a phenotype can be quantified as the average genotypic robustness of the genotypes in a genotype network [51] or as the total number of genotypes in the genotype network [8]. If in the latter case a phenotype is made up of more than a single genotype network, then the average number of genotypes per genotype network can be used to measure phenotypic robustness. Fitness robustness can be measured as the sum of the phenotypic robustness of all phenotypes that are connected via mutational events that do not yield a change in fitness.

Genotype networks provide a general framework for characterizing the neutrality of a redundant mapping, and have found application in a wide array of systems, including sulfur metabolism [39], RNA [14, 38, 41], gene regulatory networks [6, 35, 36], and field programmable gate arrays [37]. One of the primary advantages of discussing neutrality in terms of robustness is that an exact measure of neutrality can be specifically assigned to each genotype, phenotype, and fitness value [51, 52]. This allows for the assessment of the distributions of robustness, which describe the frequency with which a given robustness value is observed, at each of these three levels. Further, it allows for the quantification of the mutational biases that exist amongst genotypes, phenotypes, and fitness values. For example, the set of phenotypically-non-neutral mutations associated with a genotype need not be evenly divided amongst other genotype networks; some mutational transitions may be more likely than others.

This last point is of particular importance, as the utility of a redundant representation in evolutionary search is not only dependent upon the various distinctions of redundancy [40], but also upon the manner in which genotype space is partitioned into genotype networks and how this impacts mutational transitions amongst genotypes, phenotypes, and fitness values. For example, increasing the number of genotypes that map to a given phenotype will be of limited value if this increase does not provide mutational opportunities to discover new phenotypes. Similarly, a genetic overrepresentation of the optimal phenotype will only be advantageous if there is a corresponding increase in the number of mutational opportunities to access that phenotype. Genotype networks provide a framework for the systematic investigation of such mutational opportunities, through the characterization of phenotypically- and fitness-non-neutral genetic mutations, and thus for the concrete assessment of the potential benefits of neutrality in mutation-based

evolutionary search. Specifically, both the relationship between robustness and the ability to discover novel phenotypes, i.e., evolvability, and the relationship between robustness and the relative ease with which a phenotype is accessed by a mutation-based evolutionary process, i.e., accessibility, can be described exactly using genotype networks [8, 47, 51, 52].

The utility of neutrality in GP is a contentious topic [16]. While some studies have found no benefit [7, 44, 45], others have claimed that neutrality buffers against deleterious genetic perturbation [20, 46, 57] and reduces the risk of premature convergence through an expansion of the search space [12, 18]. However, little work has been done to explicitly characterize robustness, evolvability, and accessibility at the genotypic, phenotypic, and fitness levels, nor to describe their relationships within and between these levels [21]. However, in biological systems these relationships have been the focus of numerous theoretical [6, 10, 22, 30, 34, 35, 51, 52, 54, 55] and empirical [4, 13, 19, 23] analyses. For example, the enhanced robustness of rewired bacterial gene networks has been shown to increase cell viability in novel environments [23]. Similarly, increased robustness in the cytochrome P450 BM3 protein has been shown to increase the probability that mutants can hydroxylate novel substrates [4]. These empirical observations can be explained theoretically by considering robust phenotypes as large genotype networks, through which a population diffuses neutrally and builds up genetic diversity [22, 48]. This facilitates access to novel phenotypes through phenotypically-non-neutral mutations into adjacent genotype networks [51].

Expanding upon the work of [2], we have recently used genotype networks to describe the distributions of robustness at the level of the genotype and phenotype in a simple linear genetic programming (LGP) system used to solve a Boolean search problem [21]. This LGP system was chosen because it offers several advantages over alternative GP systems. First, the fixed-length representation is compact; the set of all genotypes is finite and computationally enumerable. Second, redundancy is intrinsic to the system; in our implementation, a total of 2^{28} genotypes map to 16 phenotypes, which in turn map to 5 fitness values. Third, there is a clear delineation between genotype, phenotype, and fitness, allowing for a full description of their interplay. By capitalizing on recent developments in the characterization of robustness, evolvability, and accessibility in RNA [8, 52], we provided a quantitative analysis of the genotype and phenotype spaces in this LGP system. We then conducted a preliminary exploration of the relationships between robustness, evolvability, accessibility, and mutation-based search, using a large ensemble of random walks.

The primary goal of our previous and current study is to describe the redundancy of this LGP system and to relate the properties of this redundancy to mutation-based evolutionary processes. To this end, we address several research questions. For example, is the redundancy uniform or non-uniform, synonymous or non-synonymous? Is neutrality present? If so, how is genotype space partitioned? How is robustness distributed amongst genotypes, phenotypes, and fitness values? What are the relationships between robustness, evolvability, and accessibility within and between each of these three levels? How do these properties relate to mutation-based evolutionary search?

Here, we embed our previous results [21] within an extended study, broadening our analysis in several ways. First, we describe the genotype networks in greater detail, providing additional topological analyses that clarify some previously unexplained observations regarding their structure. Second, we expand the scope of analysis to include robustness, evolvability, and accessibility at the fitness level. Third, we augment our evolutionary analyses to include random walks (1) that are constrained to a single genotype network and (2) that only permit mutations that maintain or improve fitness. Lastly, we use Markov chains to analytically approximate the duration and trajectory of these evolutionary processes, and we provide a mechanistic explanation for their occasional failure.

2 Methods

2.1 Linear genetic programming

In the LGP representation, an individual (or program) consists of a set of L instructions, which are structurally similar to those found in register machine languages. Each instruction is made up of an operator, a set of operands, and a return value. In the programs considered in this study, each instruction consisted of an operator drawn from the set {AND, OR, NAND, NOR}, two Boolean operands, and one Boolean return value. The inputs, operands, and return values were stored in registers with varying read/write permissions. Specifically, R_0 and R_1 were used as calculation registers that could be read and written, whereas R_2 and R_3 were used as input registers that were read-only. In this formulation, a calculation register can serve in an instruction as an operand or a return, but an input register can only be used as an operand. An example program with $L = 4$ is given below.

$$\begin{aligned} R_1 &= R_2 \text{ OR } R_3 \\ R_0 &= R_1 \text{ AND } R_2 \\ R_1 &= R_0 \text{ NAND } R_1 \\ R_0 &= R_3 \text{ NOR } R_1 \end{aligned}$$

Instructions were executed sequentially from top to bottom. Prior to program execution, the values of R_0 and R_1 were initialized to 0. After program execution, the final value in R_0 was returned as output.

2.2 Genotype, phenotype, and fitness space

To facilitate the enumeration of the entire genotype, phenotype, and fitness spaces, we considered a two-input, one-output Boolean problem instance with $L = 4$ instructions. This sequence of instructions is referred to as the *genotype*, x^g . Letting C and I denote the numbers of calculation and input registers, respectively, and O the cardinality of the operator set, there are a total of $(C \times (C + I)^2 \times O)^L$ genotypes in the LGP representation. We refer to this set of programs as the *genotype space*, Φ^g . In the system considered here ($L = 4$, $C = 2$, $I = 2$, $O = 4$),

the genotype space comprises $|\Phi^g| = 2^{28}$ unique programs and each genotype can be converted into any one of 40 neighboring genotypes with a single point mutation to one of its 16 loci.

These genotypes map to a considerably smaller set of phenotypes, which are defined by the functional relationship between the input and output registers. Specifically, the *phenotype* x^p is defined by the set of outputs observed across each of the four possible combinations of Boolean inputs. Since the outputs are also Boolean, the *phenotype space*, Φ^p comprises $|\Phi^p| = 2^4 = 16$ unique phenotypes. This genotype-phenotype map $\psi_{g \mapsto p} : \Phi^g \mapsto \Phi^p$ is thus redundant, because $|\Phi^g| > |\Phi^p|$ [40]. As an example of $\psi_{g \mapsto p}$, consider the program provided above, which yields the following truth table

$R_2[x]$	$R_3[y]$	R_0
0	0	0
0	1	0
1	0	1
1	1	0

The phenotype is the 4-bit vector in the rightmost column of the truth table, which corresponds to the function $x \text{ AND } !y$, where $!$ denotes negation.

Each of the 16 phenotypes can be assigned a fitness value x^f within the fitness space Φ^f using a mapping $\psi_{p \mapsto f} : \Phi^p \mapsto \Phi^f$ that depends upon the prescribed phenotypic target t^p . In this study, $\psi_{p \mapsto f}^{t^p}(x^p)$ is the Hamming distance between the phenotype x^p and the target t^p . We assume fitness minimization. Since the phenotypes are represented as 4-bit vectors, there are five possible fitness values and the mapping of phenotype to fitness depends upon which phenotype is chosen as the target. For example, the phenotype TRUE (i.e., $\langle 1111 \rangle$) has a fitness of 4 when the target phenotype is FALSE (i.e., $\langle 0000 \rangle$), but has an improved fitness of 1 when the target phenotype is $x \text{ OR } y$ (i.e., $\langle 0111 \rangle$).

2.3 Genotype, phenotype, and fitness networks

The redundant mapping of genotype to phenotype may generate neutrality. As mentioned in the Introduction, a convenient formalism for describing the neutrality of a redundant mapping is a *genotype network*, in which genotypes are represented as vertices and edges connect genotypes that can be interconverted via phenotypically-neutral point mutations². A genotype network $G^{x^p} = (\Gamma^{x^p}, \Theta^{x^p})$ corresponding

² Note that a phenotype may comprise multiple, independent genotype networks that cannot be reached from one another via phenotypically-neutral point mutations. Such collections of genotype networks are referred to as *genotype sets* [53]. Since it is known that the phenotypes of this LGP system each comprise a single genotype network [21], we do not present the mathematical formalisms needed to describe genotype sets.

to phenotype x^p is formally defined as a set of genotypes $\Gamma^{x^p} \subseteq \Phi^g$ and a set of edges Θ^{x^p} connecting these genotypes; an edge $\langle x^g, y^g \rangle \in \Theta^{x^p}$ if for genotype x^g an application of the mutation operator $\theta : \Phi^g \rightarrow \Phi^g$ yields a different genotype y^g of the same phenotype x^p , i.e., $\theta(x^g) = y^g$ and $\psi_{g \rightarrow p}(x^g) = \psi_{g \rightarrow p}(y^g) = x^p$. In this study, we are concerned with point mutations, which we define as a single change to an operand, operator, or return of the instruction set of a program. This point mutation is phenotypically-neutral if it does not lead to a change in phenotype (Fig. 1).

Genotype networks corresponding to different phenotypes may be connected to one another via phenotypically-non-neutral point mutations, in which case they are referred to as *adjacent*. Note that there may be many individual points of contact between adjacent genotype networks, but that each of these points corresponds to a single point mutation (Fig. 1a). Formally, two genotype networks G^{x^p} and G^{y^p} are adjacent if there exist some $x^g \in \Gamma^{x^p}$ and $y^g \in \Gamma^{y^p}$ such that $\theta(x^g) = y^g$. The set of edges that correspond to phenotypically-non-neutral point mutations between genotypes in the genotype networks of phenotypes x^p and y^p is denoted by Ω^{x^p, y^p} . By considering the adjacency of all genotype networks in the genotype space, we can construct a *phenotype network* (Fig. 2). Vertices correspond to phenotypes and are weighted according to the number of genotypes in their underlying genotype network, and edges correspond to the adjacency of genotype networks and are weighted according to the number of phenotypically-non-neutral point mutations between genotype networks. Vertices may also be assigned a fitness value, which corresponds to the phenotype's Hamming distance from a pre-specified phenotypic target t^p . Thus, we can formally define a phenotype network in three ways, depending upon whether phenotypes are assigned fitness values, and if so, whether deleterious mutations are allowed.

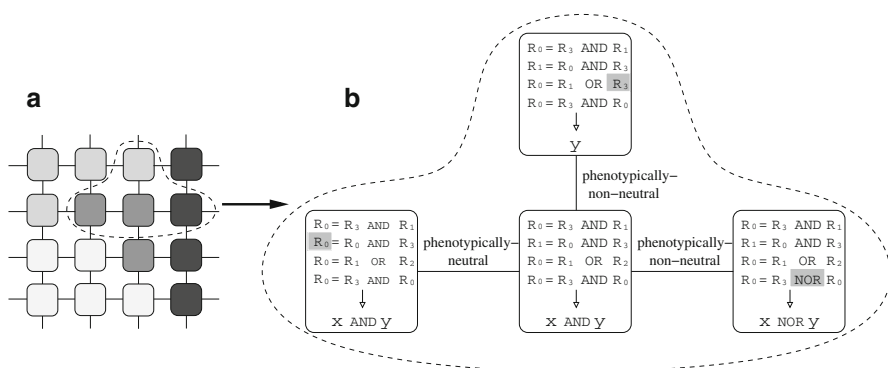


Fig. 1 **a** Schematic diagram of a subset of genotype space in linear genetic programming. Vertices correspond to genotypes, their *color* to phenotypes, and edges connect genotypes that can be interconverted via point mutations. **b** Point mutations (highlighted in gray) correspond to a single change in the instruction set and can be phenotypically-neutral or phenotypically-non-neutral, depending on whether the phenotype is preserved. For visual clarity, we only depict a small subset of the 40 potential point mutations to the 16 loci of each genotype

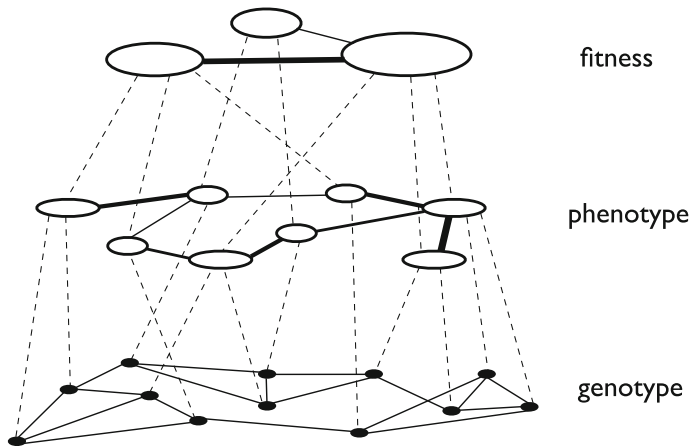


Fig. 2 Schematic diagram of the redundant mapping between genotype, phenotype, and fitness networks. The *dashed vertical lines* show that multiple vertices at a lower level can be mapped to a single vertex at a higher level. The *thickness of the solid lines* indicates the number of possible mutational transitions between vertices. *Each vertex* corresponds uniquely to a single genotype, phenotype, or fitness value

1. In the first case, phenotypes are not assigned fitness values. The phenotype network $P = (\Phi^p, \Upsilon)$ comprises the set of all phenotypes Φ^p and a set of undirected edges Υ connecting phenotypes; an edge $\langle x^p, y^p \rangle \in \Upsilon$ if $|\Omega^{x^p, y^p}| > 0$. The weight of each phenotype $x^p \in \Phi^p$ is $|\Gamma^{x^p}|$. The weight of each edge $\langle x^p, y^p \rangle \in \Upsilon$ is $|\Omega^{x^p, y^p}|$.
2. In the second case, phenotypes are assigned fitness values and deleterious mutations are allowed. The phenotype network is therefore identical to the previous case, save the fact that each phenotype now corresponds to a particular fitness value, which is determined by the phenotype-to-fitness mapping $\psi_{p \rightarrow f}^{fp}$.
3. In the third case, phenotypes are assigned fitness values, but deleterious mutations are not allowed. This corresponds to a “replace if better or equal” selection strategy [9, 24]. The phenotype network is therefore modified $P^{fp} = (\Phi^p, \Upsilon^{fp})$ such that it depends upon the target. The set of edges Υ^{fp} are now directed; an edge $\langle x^p, y^p \rangle \in \Upsilon^{fp}$ points from phenotype x^p to phenotype y^p if $|\Omega^{x^p, y^p}| > 0$ and $\psi_{p \rightarrow f}^{fp}(x^p) > \psi_{p \rightarrow f}^{fp}(y^p)$.

The mapping from phenotype to fitness $\psi_{p \rightarrow f}^{fp}$ is also redundant, with several phenotypes yielding the same fitness value. This redundancy may yield an additional layer of neutrality. By considering the connectivity of phenotypes with different fitness values, we can construct a *fitness network* (Fig. 2), where each vertex corresponds to a single fitness value and is weighted according to the sum of the sizes of the underlying phenotypes’ genotype networks. Edges correspond to the adjacency of fitness values and are weighted according to the number of fitness-non-neutral

point mutations between the genotype networks of the phenotypes that make up each fitness value. Fitness networks can be formally defined in two ways, depending upon whether deleterious mutations are allowed.

1. In the first case, deleterious mutations are allowed. The fitness network $F^{tp} = (\Phi^f, \Delta^{tp})$ corresponding to the phenotypic target t^p comprises the set of all fitness values Φ^f and a set of undirected edges Δ^{tp} connecting fitness values; an edge $\langle x^f, y^f \rangle \in \Delta^{tp}$ exists between fitness value x^f and fitness value y^f if $\exists x^g, y^g \in \Phi^g$ such that $\theta(x^g) = y^g$ and $\psi_{p \rightarrow f}^{tp}(\psi_{g \rightarrow p}(x^g)) = x^f$ and $\psi_{p \rightarrow f}^{tp}(\psi_{g \rightarrow p}(y^g)) = y^f$.
2. In the second case, deleterious mutations are not allowed. This leads to a modification of the fitness network such that the set of edges are directed. Formally, an edge $\langle x^f, y^f \rangle \in \Delta^{tp}$ points from fitness value x^f to fitness value y^f if $\exists x^g, y^g \in \Phi^g$ such that $\theta(x^g) = y^g$ and $\psi_{p \rightarrow f}^{tp}(\psi_{g \rightarrow p}(x^g)) = x^f$ and $\psi_{p \rightarrow f}^{tp}(\psi_{g \rightarrow p}(y^g)) = y^f$ and $x^f > y^f$.

In both cases, the set of edges between the genotype networks of the phenotypes with fitness values x^f and y^f is denoted by Ω^{x^f, y^f} . The weight of each fitness value $x^f \in \Phi^f$ is $\sum_{\{x^p | \psi_{p \rightarrow f}^{tp}(x^p) = x^f\}} |\Gamma^{x^p}|$. The weight of each edge $\langle x^f, y^f \rangle \in \Delta^{tp}$ is $|\Omega^{x^f, y^f}|$.

2.4 Observable quantities

To characterize the genotypic, phenotypic, and fitness spaces of this LGP system, we consider the topological measures defined below. In addition to describing how these quantities relate to one another, both within and between levels, we will also consider their relationship with simple, mutation-based evolutionary processes.

2.4.1 Robustness

We use robustness to quantify the degree of neutrality associated with each genotype, phenotype, and fitness value. Specifically, we define *genotypic robustness* R^g as [52]

$$R^g(x^g) = k^{x^g}/40, \quad (1)$$

where k^{x^g} is the number of connections genotype x^g possesses in the genotype network and 40 is the total number of possible point mutations. Genotypic robustness is thus the fraction of the total number of possible point mutations to a given genotype that are phenotypically-neutral. We define *phenotypic robustness* R^p as the number of genotypes in the phenotype's underlying genotype network,

$$R^p(x^p) = |\Gamma^{x^p}|. \quad (2)$$

This is the number of genotypes that yield the same phenotype and that are connected via a series of phenotypically-neutral point mutations. We define *fitness robustness* R^f as the sum of the phenotypic robustnesses of all phenotypes with a given fitness value,

$$R^f(x^f) = \sum_{\{x^p | \psi_{p \rightarrow f}(x^p) = x^f\}} R^p(x^p). \quad (3)$$

2.4.2 Evolvability

Several definitions of evolvability have been put forth [26, 52, 54]. Here, we focus on those definitions that can be expressed in terms of the properties of genotype networks. We define *genotypic evolvability* E^g of a genotype x^g with phenotype x^p as the proportion of the total number of possible phenotypes that can be reached via individual, phenotypically-non-neutral point mutations to genotype x^g (i.e., all genotypes that are of edit distance 1 from x^g) [52]

$$E^g(x^g) = |\{y^p | \theta(x^g) = y^g, \psi_{g \rightarrow p}(y^g) = y^p, x^p \neq y^p\}|/15 \quad (4)$$

For *phenotypic evolvability*, we consider two measures. The first measure, E_1^p , is simply the proportion of the total number of possible phenotypes that are adjacent to a given phenotype (i.e., via phenotypically-non-neutral point mutations to genotypes in the underlying genotype network) [52],

$$E_1^p(x^p) = k^{x^p}/15 \quad (5)$$

where k^{x^p} is the number of edges emanating from phenotype x^p in the phenotype network. In Eqs. (4) and (5), the denominator is the total number of possible adjacent phenotypes. The second measure, E_2^p , provides a more nuanced analysis of the potential to mutate from one phenotype to another [8]. Letting

$$f_{x^p y^p}^p = \begin{cases} \frac{|\Omega^{x^p y^p}|}{\sum_{z \neq y} |\Omega^{x^p z^p}|}, & \text{if } x^p \neq y^p \\ 0, & \text{if } x^p = y^p \end{cases} \quad (6)$$

denote the proportion of phenotypically-non-neutral point mutations to genotypes of phenotype x^p that result in genotypes of phenotype y^p , we define the evolvability E_2^p of phenotype x^p as

$$E_2^p(x^p) = \begin{cases} 0, & \text{if } f_{x^p y^p}^p = 0 \forall y^p \\ 1 - \sum_{y^p} (f_{x^p y^p}^p)^2 & \text{otherwise.} \end{cases} \quad (7)$$

This corresponds to the probability that two randomly chosen phenotypically-non-neutral point mutations to genotypes of phenotype x^p result in genotypes with distinct phenotypes. Thus, this measure takes on a value of zero if phenotype x^p can only mutate into one other phenotype. More generally, this measure takes on low values when a phenotype is adjacent to only a few other phenotypes and its phenotypically-non-neutral mutations are biased toward a subset of these phenotypes. It takes on high values when a phenotype is adjacent to many other phenotypes and its phenotypically-non-neutral mutations are uniformly divided amongst these phenotypes.

These two measures can be adapted to measure *fitness evolvability*. The first measure E_1^f corresponds to the proportion of all fitness values possibly adjacent to fitness value x^f

$$E_1^f(x^f) = k^{x^f}/4, \quad (8)$$

where k^{x^f} is the number of connections emanating from fitness value x^f in the fitness network and 4 is the total number of possible adjacent fitness values. The second measure uses

$$f_{x^f y^f}^f = \begin{cases} \frac{|\Omega^{x^f, y^f}|}{\sum_{z \neq y^f} |\Omega^{x^f, z^f}|}, & \text{if } x^f \neq y^f \\ 0, & \text{if } x^f = y^f \end{cases} \quad (9)$$

to measure fitness evolvability E_2^f as

$$E_2^f(x^f) = \begin{cases} 0, & \text{if } f_{x^f y^f}^f = 0 \forall y^f \\ 1 - \sum_{y^f} (f_{x^f y^f}^f)^2 & \text{otherwise.} \end{cases} \quad (10)$$

2.4.3 Accessibility

In addition to measuring phenotypic evolvability E_2^p , which describes the uniformity of phenotypically-non-neutral mutations emanating from phenotype x^p , we also measure *phenotypic accessibility* [8],

$$A^p(x^p) = \sum_{y^p} f_{y^p x^p}^p, \quad (11)$$

which represents the propensity to mutate into phenotype x^p . This measure takes on high values if a phenotype is relatively easy to access from other phenotypes, and low values otherwise³. For an alternative formulation of this quantity see [27, 28].

The analogous definition of *fitness accessibility* is

$$A^f(x^f) = \sum_{y^f} f_{y^f x^f}^f, \quad (12)$$

which represents the propensity to mutate into fitness value x^f . Note that when deleterious mutations are not allowed, the worst fitness value has $A^f(x^f = 4) = 0$.

2.4.4 Distance and diversity

The distance between two genotypes x^g and y^g is calculated as

$$D(x^g, y^g) = \frac{1}{16} \sum_{i=1}^{16} \delta(x_i^g, y_i^g), \quad (13)$$

where $\delta(x_i^g, y_i^g) = 1$ if genotypes x^g and y^g differ at location i and $\delta(x_i^g, y_i^g) = 0$ otherwise. The summation is taken across all 16 loci and then normalized.

³ Note that genotypic accessibility is not a useful concept. This is because it is equivalent to the total number of possible point mutations to a genotype. Specifically, if we used $f_{x^g y^g}^g$ to denote the fraction of point mutations to genotype x^g that result in genotype y^g , then $\sum_{y^g} f_{y^g x^g}^g = 40 \forall x^g \in \Phi^g$.

The diversity of the sets of phenotypes Λ^{x^g} and Λ^{y^g} that are accessible within the 1-neighborhood of two genotypes x^g and y^g of the same phenotype x^p is calculated as [6]

$$F(\Lambda^{x^g}, \Lambda^{y^g}) = |\overline{\{\Lambda^{x^g} \cap \Lambda^{y^g}\}}|/15, \quad (14)$$

where $\overline{\{*\}}$ denotes set complement. The denominator reflects the extreme case where one genotype can access all of the 15 possibly adjacent phenotypes and the other can access none (e.g., $\Lambda^{x^g} = \Phi^p \setminus \{x^p\}$, $\Lambda^{y^g} = \emptyset$). If F is small, the two genotypes are adjacent to similar sets of genotype networks. If F is large, the two genotypes have mutational access to diverse sets of genotype networks.

2.4.5 Coreness

The genotype networks considered in this study are too large to visualize directly. However, we can gain further insight into their structure by describing the distributions of certain vertex-level properties. One such property is vertex degree, which we use to calculate genotypic robustness. Another property is *coreness*, which is an integer index k that defines the position of a vertex as belonging to one of several non-overlapping k -shells [33]. Each k -shell is defined as a subset of vertices in which each vertex is connected to at least k other vertices. Thus, vertices with large k are close to the innermost core of the network, and vertices with small k are nearer the periphery.

3 Results

3.1 Statistical characteristics of genotype, phenotype, and fitness spaces

To investigate the genotype, phenotype, and fitness spaces of the two-input, one output LGP system of $L = 4$ instructions, we exhaustively enumerated all 2^{28} genotypes, which allowed for a full characterization of their mutational connectivities. We present our analysis of these spaces incrementally, beginning at the level of the genotype and ending with a description of the interplay between genotype, phenotype, and fitness. We use the measures provided in Sect. 2.4 to describe these spaces and we often illustrate the relationships between these quantities using correlations, which we summarize with Pearson's correlation coefficient and a standard permutation test of statistical significance.

Our goal is to use these measures to address the following research questions, among others. What is the relationship between evolvability and robustness at the genotypic level? Is this relationship different at the phenotypic or fitness levels? Are genotype networks confined to specific regions of genotype space or do they extend throughout its entirety? Does the location of a genotype on a genotype network impact the set of phenotypes it can access via phenotypically-non-neutral mutation? How does the choice of phenotypic target impact the structure of the phenotype network and fitness network?

3.1.1 Genotype space

Genotype space was partitioned into 16 independent genotype networks. These genotype networks ranged in size from a minimum of 24,832 genotypes to a maximum of 60,393,728 genotypes (ranging in size from $\ll 1\%$ to 23% of genotype space, respectively).

In Fig. 3, we depict several properties of the genotype network that corresponds to the representative phenotype !x AND y. In this particular genotype network, as well as all others in this LGP system, the distribution of genotypic evolvability is unimodal (Fig. 3a), while the distribution of genotypic robustness is bimodal (Fig. 3b). These quantities exhibit a slight, but highly significant, inverse relationship, such that genotypes of greater robustness are generally less evolvable ($R^2 = 0.01$, $p \ll 0.01$, Fig. 3c).

The coreness of a genotype is shown as a function of its robustness in Fig. 3d. The data are positively correlated ($R^2 = 0.79$, $p \ll 0.01$) and fall into two discrete clusters, suggesting that the genotype network consists of a single dense core of highly robust genotypes and a periphery of less robust genotypes. The clear delineation of the two clusters explains the bimodality in the distribution of genotypic robustness (Fig. 3b).

Figure 3e shows the distribution of genotypic distance between randomly sampled genotypes in this genotype network. The distribution is unimodal, with an average distance that falls within one standard deviation of the mean of the corresponding null distribution (vertical dashed lines). This indicates that the genotype network is not restricted to a specific region of genotype space, but instead extends broadly into distant regions of genotype space. The average genotypic distance grows logarithmically from a minimum of 0.62 to a maximum of 0.68 as the size of the genotype network increases (Fig. 3e, inset, $R^2 = 0.91$, $p \ll 0.01$), but never falls outside the bounds of the null distribution.

To assess the implications of such expansive genotype networks, we calculated the diversity of the genotype networks adjacent to each randomly chosen pair of genotypes. The diversity of adjacent genotype networks is shown as a function of genotypic distance in Fig. 3f. In general, the diversity of adjacent genotype networks increases as the distance between two genotypes increases (the non-monotonicity of the trend is attributable to undersampling at the tails of the distribution, cf., Fig. 3e). This indicates that a genotype's position in genotype space has a strong influence on the genotype networks that surround it.

3.1.2 Phenotype space

Each of the 16 genotype networks in this system correspond uniquely to a single phenotype. As such, any two genotypes that yield the same phenotype are connected through a series of phenotypically-neutral point mutations. The phenotype network of the mutational transitions between these 16 genotype networks is depicted in Fig. 4. The network is fully connected, such that any phenotype can be reached directly from any other. However, the number of phenotypically-non-neutral point mutations between phenotypes, depicted by edge width, is heterogeneous. Some

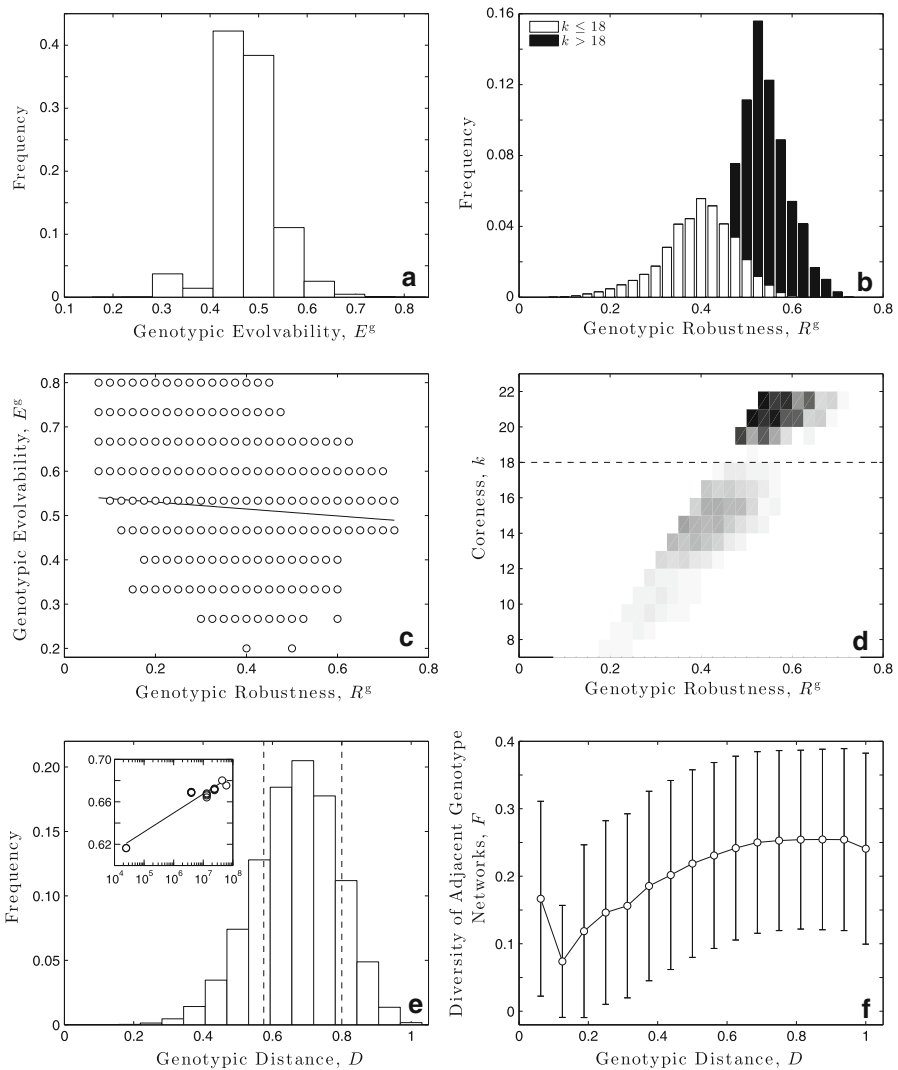


Fig. 3 Properties of genotypes within the phenotype !x AND y. Distributions of **a** genotypic evolvability E^g and **b** genotypic robustness R^g for all ≈ 4 million genotypes. **c** Genotypic evolvability E^g as a function of genotypic robustness R^g . The solid line represents the best linear fit to the data and is provided as a guide for the eye. **(d)** Coreness k as a function of genotypic robustness R^g . Data are linearly binned, with darker bin shades indicating higher frequency. Note the clear delineation between genotypes with coreness $k \leq 18$ and $k > 18$ (dashed horizontal line). This arbitrary distinction is used to color the bars in **b**, indicating that highly robust genotypes reside in the core of the genotype network. **e** Distribution of the genotypic distance D between 200,000 randomly sampled pairs of genotypes. The dashed vertical lines represent one standard deviation from the mean of the corresponding null distribution, which was determined by sampling pairs of genotypes at random from the entire genotype space (i.e., without regard to phenotype) and calculating the genotypic distance between these pairs. The inset depicts the mean genotypic distance for all 16 genotype networks as a function of their size. The solid line represents the best logarithmic fit to the data and is provided as a guide for the eye. **(f)** The diversity of adjacent genotype networks F is shown as a function of genotypic distance D

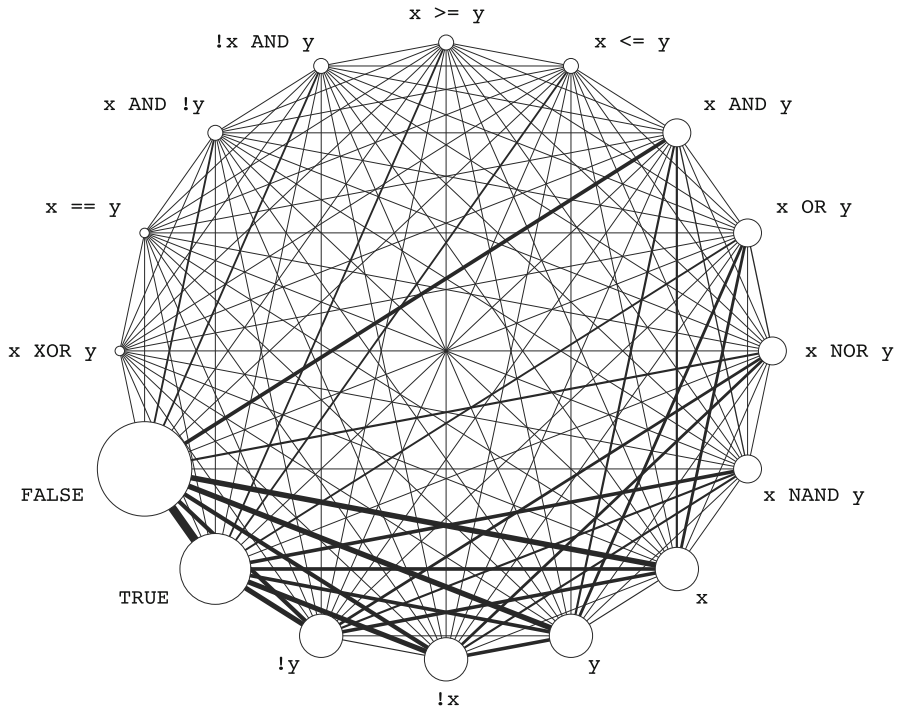


Fig. 4 Phenotype network for linear genetic programming with two inputs, one output, and four instructions. Each vertex comprises a genotype network, as depicted schematically in Fig. 1, and thus vertex size corresponds to phenotypic robustness. *Edge width* denotes the number of phenotypically-non-neutral point mutations between two phenotypes, and is normalized by the total number of phenotypically-non-neutral point mutations between all pairs of phenotypes. Phenotypes are labeled according to their functional relationship between input and output, where x and y denote the inputs stored in registers R_2 and R_3 , respectively

phenotypes are mutationally biased toward a small subset of phenotypes (e.g., Fig. 4, $x \text{ AND } y$), while others mutate nearly uniformly to all other phenotypes (e.g., Fig. 4, $x == y$). Note that the edges are undirected, because in the absence of fitness there are no deleterious mutations. All mutational events between phenotypes are reversible and therefore symmetric. Phenotypic robustness is denoted by vertex size, and the variety of vertex sizes mirrors the heterogeneous distribution of the sizes of the underlying genotype networks. Note that the phenotype **FALSE** is larger than the phenotype **TRUE**, despite the inherent symmetry of this LGP system. This occurs because the output register R_0 is initialized to 0. Therefore, if a program does not modify its output register, its default phenotype is **FALSE**.

The means of the distributions of genotypic evolvability and robustness vary as a function of phenotypic robustness (Fig. 5a, b). Specifically, average genotypic evolvability decreases logarithmically as a function of phenotypic robustness (Fig. 5a, $R^2 = 0.95$, $p \ll 0.01$). This intuitive observation implies that within robust phenotypes, most mutations are phenotypically-neutral and do not allow access to adjacent phenotypes. It follows that the individual genotypes that make up robust

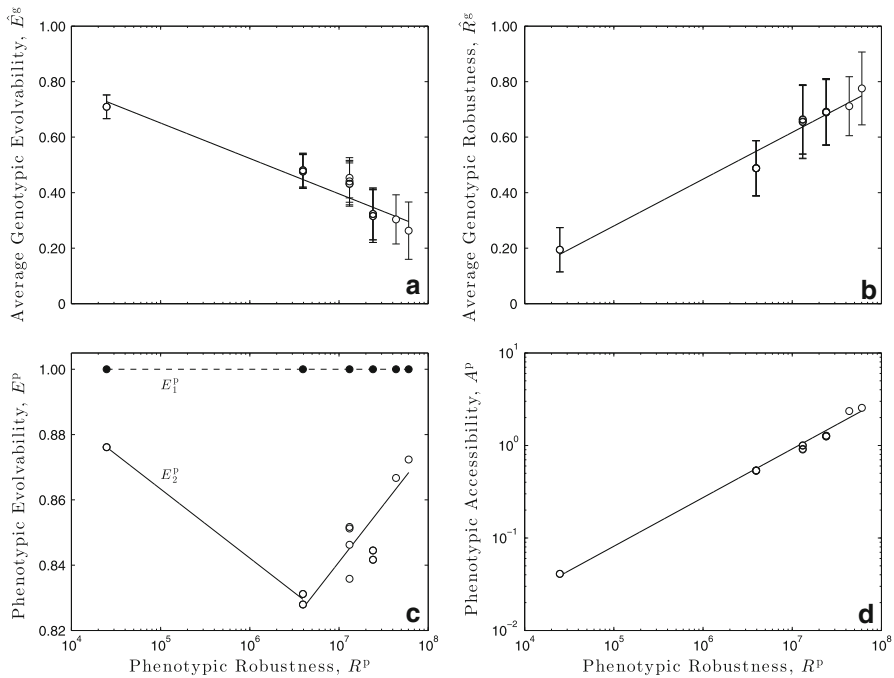


Fig. 5 Properties of phenotype space and their relation to genotype space. Average genotypic **a** evolvability \hat{E}_g^g and **b** robustness \hat{R}_g^g and phenotypic, **c** evolvability E_1^P , E_2^P and **d** accessibility A^P as a function of phenotypic robustness R^P . The data in **(a, b)** correspond to the average of all genotypes within a given phenotype and error bars denote their standard deviation. The *solid lines* correspond to the best **(a, b)** logarithmic, **c** piecewise logarithmic, and **d** power-law fit to the data, and are provided as a guide for the eye

phenotypes are collectively more robust. Indeed, we observe that the average genotypic robustness increases logarithmically as a function of phenotypic robustness (Fig. 5b, $R^2 = 0.98$, $p \ll 0.01$).

The relationship between phenotypic evolvability and phenotypic robustness is less intuitive. Because the phenotype network is fully connected, all phenotypes are equally and maximally evolvable according to E_1^P (filled circles, Fig. 5c). In contrast, when mutational biases are taken into account with E_2^P , phenotypic evolvability exhibits a nonlinear relationship with phenotypic robustness (open circles, Fig. 5c). Phenotypic evolvability is lowest for phenotypes of intermediate robustness (xAND!y, !xANDy), and then increases logarithmically with increasing phenotypic robustness ($R^2 = 0.87$, $p = 0.02$). The relationship is made non-monotonic by the high evolvability of the least robust phenotypes (xXORy, x == y).

Phenotypic accessibility increases monotonically as a function of phenotypic robustness, following the power-law $A^P \propto (R^P)^{1/2}$ (Fig. 5d, $R^2 = 0.99$, $p \ll 0.01$). This implies that random mutations are more likely to lead to robust than to non-robust phenotypes. Taken together, these results suggest that the most robust phenotypes are both easy to find (Fig. 5d) and highly evolvable (Fig. 5c), with the

exception of the least robust phenotypes, which are simultaneously the least accessible and the most evolvable of any of the phenotypes in this system.

3.1.3 Fitness space

Due to the inherent symmetry of certain pairs of phenotypes (e.g., $x \text{ AND } !y$ and $!x \text{ AND } y$), there are 11 unique mappings of phenotype to fitness. Further, the fitness networks can be constructed under two different assumptions regarding deleterious mutations. This results in a total of 22 distinct fitness networks.

We first consider the case where deleterious mutations are allowed. In Fig. 6a, b we depict two fitness networks, for the phenotypic targets **TRUE** and $x == y$, respectively. The distributions of fitness robustness (vertex size) and the number of fitness-non-neutral mutations (edge width) are heterogeneous and vary between the 11 fitness networks. The set of phenotypes that make up each fitness value also varies depending on the phenotypic target (e.g., compare vertices of the same shade

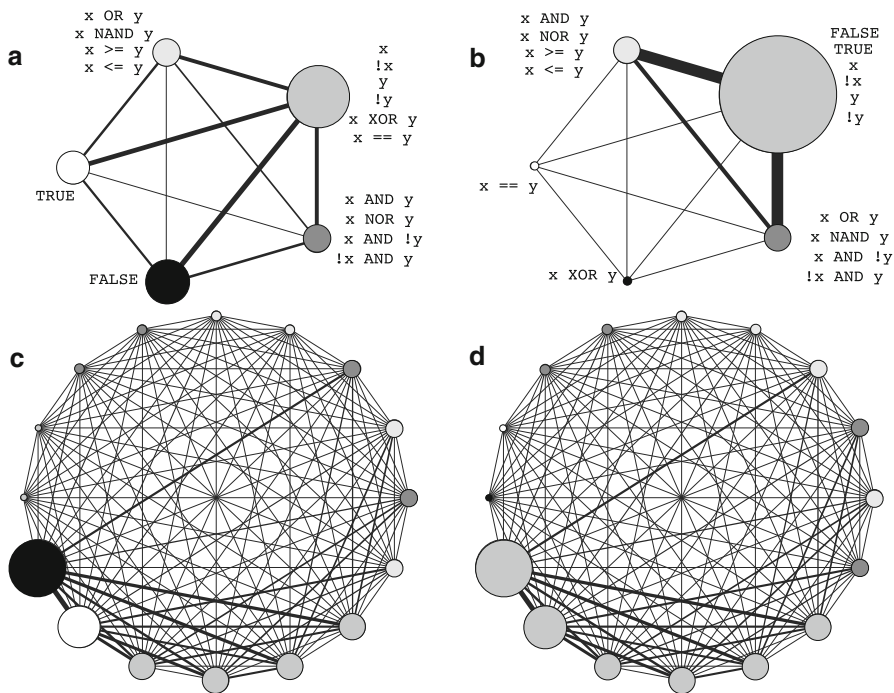


Fig. 6 Fitness networks for the phenotypic targets **a** **TRUE** and **b** $x == y$, when deleterious mutations are allowed. Each vertex represents a fitness value. Vertex annotation reflects all of the phenotypes within each fitness value and vertex size corresponds to fitness robustness. Vertex color is used to depict the fitness value, which shifts from black to white as the fitness value improves. The width of the undirected edges between two vertices corresponds to the number of fitness-non-neutral point mutations, normalized by the total number of fitness-non-neutral point mutations between all fitness values. Phenotype networks for the phenotypic targets **c** **TRUE** and **d** $x == y$. Vertex color is the same as **a**, **b**. Vertex annotation, size, and edge width are as in Fig. 4

between Fig. 6a, b), but the number of phenotypes per fitness value always remains the same.

The structure of the underlying phenotype network (Fig. 6c, d) is identical to that of Fig. 4, except that each phenotype now possesses a fitness value, which varies depending upon the phenotypic target (e.g., compare Fig. 6c, d). The relationship between the robustness, evolvability, and accessibility of a phenotype therefore does not vary between phenotypic targets.

Considering all 11 phenotypic targets simultaneously, the means of the distributions of phenotypic robustness within each fitness value are positively correlated with fitness robustness (data not shown; $R^2 = 0.85$, $p \ll 0.01$), as are the means of the distributions of phenotypic accessibility (data not shown; $R^2 = 0.81$, $p \ll 0.01$). The means of the distributions of phenotypic evolvability E_2^p exhibit a nonlinear relationship with fitness robustness (data not shown), akin to the trend depicted in Fig. 5c. Robustness, accessibility, and evolvability each therefore exhibit functional relationships between the phenotypic and fitness levels.

At the level of the fitness network, the correlation between fitness evolvability and fitness robustness is weak ($R^2 = 0.06$, $p = 0.03$; Fig. 7a). However, fitness accessibility exhibits a strong positive correlation with fitness robustness (Fig. 7b), again increasing according to the power-law $A^f \propto (R^f)^{1/2}$ ($R^2 = 0.90$, $p \ll 0.01$). Thus, the most robust fitness values are also the most accessible, an intuitive result given the positive correlation between fitness robustness and average phenotypic accessibility.

Next, we consider the case where deleterious mutations are not allowed. In Fig. 8a, b we again depict the fitness networks for the phenotypic targets TRUE and $x = y$, respectively. The sets of phenotypes that make up each fitness value are the same as in Fig. 6a, b. However, the edges are now directed and the fitness networks are weakly connected. This implies that a directed path may not exist between two vertices, reflecting the fact that mutational transitions between phenotypes are only permitted if those transitions are beneficial or fitness-neutral.

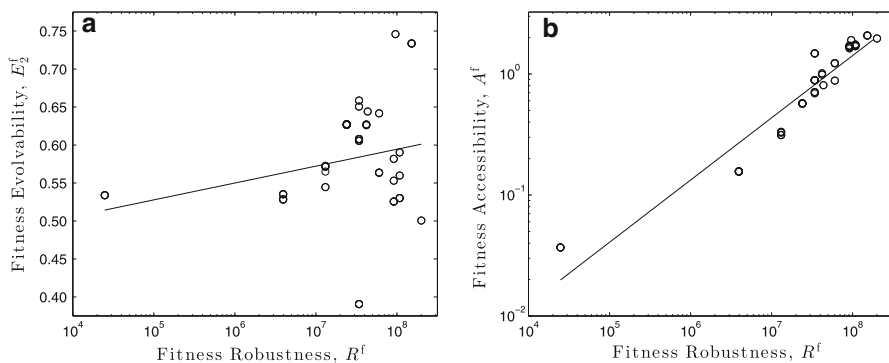


Fig. 7 Properties of fitness space when deleterious mutations are allowed. Fitness **a** evolvability E_2^f and **b** accessibility A^f as a function of fitness robustness R^f . The *solid lines* correspond to the best **a** logarithmic fit to the data and **b** the power-law $A^f \propto (R^f)^{1/2}$. Both are provided as a guide for the eye

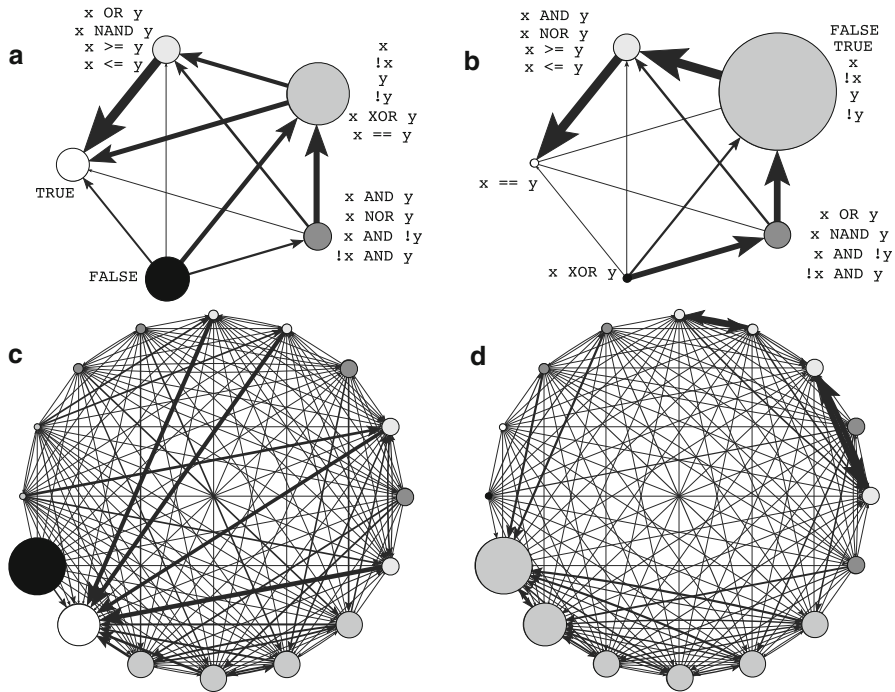


Fig. 8 Fitness networks for the phenotypic targets **a** TRUE and **b** $x == y$, when deleterious mutations are not allowed. Vertex color, size, and annotation are as in Fig. 6a, b. The width of the directed edges between two vertices corresponds to the number of fitness-non-neutral point mutations, normalized by the total number of fitness-non-neutral point mutations emanating from each fitness value. Note that the worst fitness value (black vertex) only has edges pointing away from it and the best fitness value (white vertex) only has edges pointing into it. Phenotype networks for the phenotypic targets **c** TRUE and **d** $x == y$. Vertex color, size, and annotation are as in Fig. 6c, d. The width of the directed edges between two vertices corresponds to the number of phenotypically-non-neutral point mutations, normalized by the total number of phenotypically-non-neutral point mutations emanating from each phenotype. Note that edges only point toward phenotypes of equal or better fitness (same or lighter color)

Prohibiting deleterious mutations affects the underlying phenotype network (Fig. 8c, d), leading to several fundamental changes in its structure. First, the number of phenotypically-non-neutral mutational events between phenotypes is no longer symmetric; the edges are now directed. Second, the network transforms from strongly connected to weakly connected, meaning that some phenotypes are left unreachable from some others. Third, the edge sets vary markedly between the 11 unique phenotype networks (e.g., compare Fig. 8c, d). Therefore, the relationship between the robustness, evolvability, and accessibility of phenotypes varies between phenotypic targets.

Indeed, any correlation that was previously observed between phenotypic robustness and phenotypic evolvability (Fig. 5c) is now lost, as measured using either E_2^p or E_3^p (data not shown; E_1^p : $3.1 \times 10^{-5} \leq R^2 \leq 2.7 \times 10^{-2}$, $p > 0.55$; E_2^p : $1.4 \times 10^{-5} \leq R^2 \leq 1.7 \times 10^{-1}$, $p > 0.11$, for all 11 phenotype networks). However, the correlation between phenotypic robustness and phenotypic

accessibility remains both positive and significant (data not shown; $0.48 \leq R^2 \leq 0.75$, $p \ll 0.01$ for all 11 phenotype networks). Thus, in the absence of deleterious mutations, the most robust phenotypes generally remain the easiest to find, but are no longer the most evolvable.

Since the prohibition of deleterious mutations does not impact phenotypic robustness, the correlation between fitness robustness and the means of the distributions of phenotypic robustness within each fitness value is identical to the previous case where deleterious mutations were allowed (data not shown; $R^2 = 0.85$, $p \ll 0.01$). However, prohibiting deleterious mutations does impact phenotypic evolvability and phenotypic accessibility. Specifically, the correlation between fitness robustness and the means of the distributions of phenotypic accessibility is lost (data not shown; $R^2 = 0.09$, $p = 0.02$), as is the correlation between fitness robustness and the means of the distributions of phenotypic evolvability \bar{E}_2^p (data not shown; $R^2 = 0.07$, $p = 0.01$).

At the level of the fitness network, there is no correlation between fitness evolvability and fitness robustness (data not shown; E_1^f : $R^2 = 0$, $p = 1.00$; E_2^f : $R^2 = 1.8 \times 10^{-4}$, $p = 0.91$) nor between fitness accessibility and fitness robustness (data not shown; $R^2 = 2.3 \times 10^{-3}$, $p = 0.71$). The robustness of a fitness value therefore does not affect the ease with which it is identified, an observation that stems from the lack of correlation between fitness robustness and the average accessibility of the phenotypes that make up a fitness value.

3.2 Random walks and hill climbing

To understand how the structure of genotype, phenotype, and fitness networks influence evolutionary search, we conduct four interrelated analyses. Each is a highly stylized abstraction of an evolutionary process, in which we consider the behavior of only a single individual, which is subject to mutation. These analyses are also introduced incrementally. We begin with random walks that explore a single genotype network, and end with a hill climber that concurrently traverses genotype, phenotype, and fitness space.

Our goal is to relate the dynamical properties of mutation-based search with the structural properties of genotype, phenotype, and fitness networks presented in Sect. 3.1. We address the following questions, among others: How does the robustness of a genotype influence the frequency with which a random walk encounters that genotype? Is the waiting time of a random walk to reach a target phenotype correlated with the evolvability of the phenotype in which the walk began? Can we predict waiting times using Markov chains?

3.2.1 Random walks through genotype space

In our first analysis, we consider a random walk in the genotype network of the representative phenotype !x AND y. Each step in the random walk corresponds to a single point mutation. We record the robustness of the genotype encountered in each step, and use this to calculate the *visit frequency*, which is the distribution of the

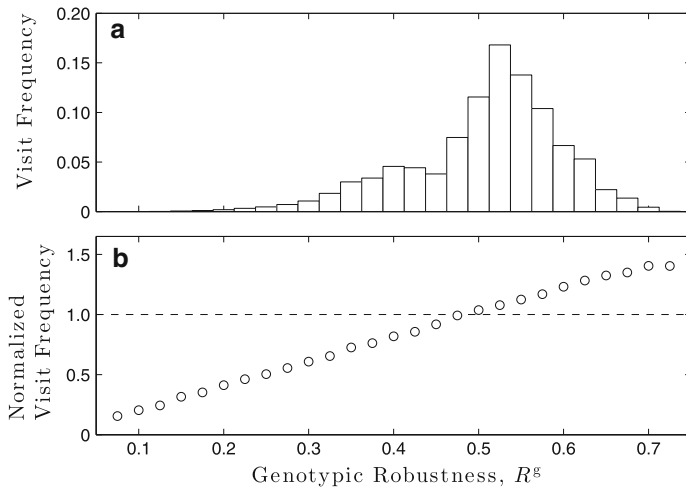


Fig. 9 A random walk in the genotype network of phenotype !x AND y. **a** Visit frequency, defined as the proportion of steps in a random walk that are spent at a genotype of a given robustness. The bin centers are the same as in Fig. 3b. **b** Visit frequency normalized by the frequency with which a given genotypic robustness value is observed in the genotype network (i.e., the distribution in **a** divided by the distribution in Fig. 3b). The horizontal dashed line indicates a visit frequency that is exactly proportional to the frequency with which a given genotypic robustness value is observed in the genotype network

proportion of steps spent at each genotypic robustness value. To ensure sufficient sampling, the number of steps in the random walk is set to the number of genotypes in the genotype network (≈ 4 million).

The visit frequency is depicted in Fig. 9a. The distribution is bimodal, akin to the distribution of genotypic robustness presented in Fig. 3b. However, dividing the former distribution (Fig. 9a) by the latter (Fig. 3b) reveals that the two distributions are in fact distinct (Fig. 9b). Genotypes are not visited uniformly, but rather in proportion to their robustness. Thus, genotypes of high robustness are visited more often, and genotypes of low robustness less often, than would be expected in a random sampling of genotypes from the genotype network. This comes with the caveat that low-evolvability genotypes are visited more often than high-evolvability genotypes (Fig. 3c).

3.2.2 Random walks through genotype and phenotype space

In our second analysis, we consider all of the 16 interconnected genotype networks, using random walks to explore both genotype and phenotype space.

For each of the 16×15 possible combinations of pairs of unique phenotypes, we designate one phenotype as a source and the other as a target. We then perform 1,000 random walks, starting from a randomly chosen genotype in the source phenotype and ending when the random walk reaches any genotype in the target phenotype. We record the average number of steps required to get from one phenotype to another, which we refer to as the mean waiting time, T_w .

In Fig. 10, we depict the mean waiting time of a random walk as a function of the source phenotype's evolvability (Fig. 10a) and the target phenotype's accessibility (Fig. 10b). The mean waiting time is independent of the evolvability of the source phenotype ($R^2 = 0.001$, $p = 0.62$), due to the Markovian nature of the random walk. However, it is strongly correlated with the accessibility of the target phenotype ($R^2 = 0.99$, $p \ll 0.01$). Specifically, the mean waiting time decreases as a function of target accessibility according to the power-law $T_W \propto (A^P)^{-5/3}$. Thus, accessible phenotypes are found more rapidly by random mutation than less accessible phenotypes. As highly accessible phenotypes are also highly robust (Fig. 5d), random mutation leads to robust phenotypes. We note that the waiting time for random search to identify a target phenotype is inversely proportional to the phenotypic robustness of the target.

The Markovian nature of the random walk suggests that an analytical determination of the expected waiting time between source and target phenotypes may be possible (“Appendix 1: Markov chains to determine the mean first passage time”) [17]. In brief, the analytical treatment uses the phenotype network (Fig. 4) as the transition matrix of a Markov chain, and designates the target phenotype as an absorbing state. After some algebraic rearrangement and a single matrix inversion, the waiting time is easily calculated. A fundamental assumption of this analysis is that the phenotype network accurately encodes the mutational transitions between phenotypes, despite the fact that the actual mutational transitions are occurring at the level of genotypes.

The analytically determined and empirically observed waiting times are strongly correlated (data not shown; $R^2 = 0.99$, $p \ll 0.001$). However, the average residual between analysis and observation is 126 steps, which constitutes approximately 15 % of the empirically observed waiting time. Such a large discrepancy suggests that the phenotype network does not actually provide an accurate description of the mutational transitions between phenotypes, as will be revealed in the subsequent section.

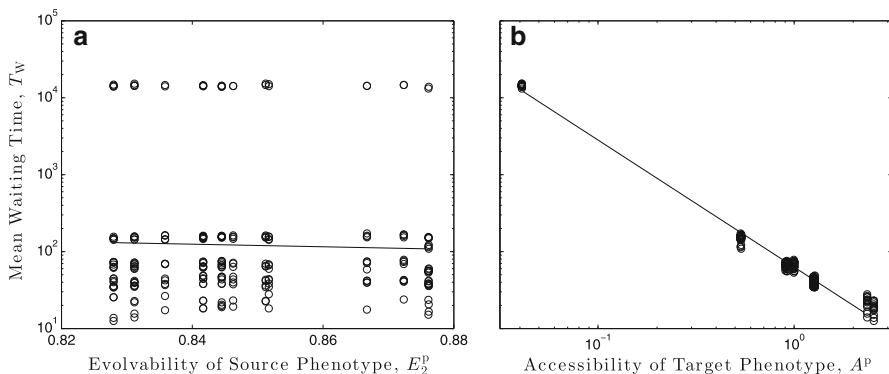


Fig. 10 Random walks in genotype and phenotype space. Mean waiting time T_W as a function of **a** the source phenotype's evolvability E_2^P and **b** the target phenotype's accessibility A^P . The solid lines correspond to the best **a** exponential and **b** power-law fit to the data, and are provided as a guide for the eye

3.2.3 Random walks through genotype, phenotype, and fitness space

In our third analysis, we consider the case where each phenotype is assigned a fitness value and deleterious mutations are allowed. We therefore use the same random walk data as in Sect. 3.2.2 to explore the relationships between the properties of fitness values and the number of steps required to get from a source phenotype to a target phenotype.

We depict the mean waiting time as a function of the source fitness value's evolvability in Fig. 11a and as a function of the target fitness value's accessibility in Fig. 11b. The mean waiting time decreases exponentially as the evolvability of the source fitness value increases ($R^2 = 0.61$, $p \ll 0.01$), in contrast to the analogous observation made at the level of the phenotype (Fig. 10a). Mean waiting time decreases according to a power-law as the accessibility of the target fitness value increases ($R^2 = 0.94$, $p \ll 0.01$). The relationship between mean waiting time and accessibility can be approximated by the function $T_w \propto (A^f)^{-2}$.

Note that since deleterious mutations are allowed, the phenotype networks have not been structurally modified (cf., Fig. 6c, d). The previously observed relationships between mean waiting time, phenotypic evolvability, and phenotypic accessibility (Fig. 10) therefore remain intact. Further, the results of the Markov chain analysis presented in Sect. 3.2.2 are identical to those found in this section.

3.2.4 Hill climbing through genotype, phenotype, and fitness space

In our fourth analysis, we consider the case where each phenotype is assigned a fitness value, but deleterious mutations are not allowed. We employ an ensemble of hill climbers that sample potential movements at random, but only accept steps that maintain or improve fitness. We again investigate each of the 16×15 possible combinations of unique source and target phenotypes. For each combination of source and target phenotype, we perform 100,000 hill climbing simulations, starting from a randomly chosen genotype in the source phenotype and ending when the walk reaches any genotype in the target phenotype. The duration of this trajectory is referred to as the mean adaptation time, T_A .

We depict the mean adaptation time as a function of the source fitness value's evolvability in Fig. 11c and as a function of the target fitness value's accessibility in Fig. 11d. Mean adaptation time and the evolvability of the source fitness value are uncorrelated ($R^2 = 0.02$, $p = 0.31$), contrasting with the previous case where deleterious mutations were allowed. Note that some fitness values have $E_2^f = 0$ because they only point to the optimal fitness value (cf., Fig. 8a, b). Mean adaptation time remains correlated with the accessibility of the target fitness value ($R^2 = 0.64$, $p \ll 0.01$).

The directed phenotype networks (e.g., Fig. 8c, d) can be used as transition matrices in Markov chains to analytically determine the expected adaptation time between source and target phenotypes (“[Appendix 1: Markov chains to determine the mean first passage time](#)”) [17]. We again find a strong correlation between the analytical and empirical results ($R^2 = 0.97$, $p \ll 0.01$), but a relatively large discrepancy in the average residual difference in adaptation time (112 steps). Similar

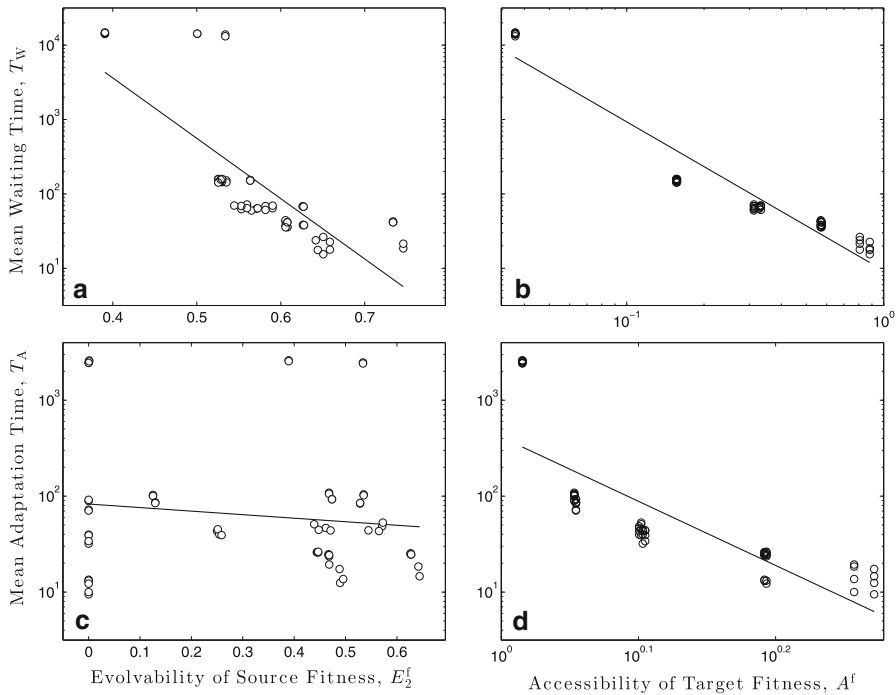


Fig. 11 Random walks and hill climbing in genotype, phenotype, and fitness space. **a, b** Mean waiting time T_W and **c, d** mean adaptation time T_A as a function of **a, c** the evolvability of the source fitness value E_2^f and **b, d** the accessibility of the target fitness value A^f , when deleterious mutations are **a, b** and are not **c, d** allowed. The *solid lines* correspond to the best **a, c** exponential and **b, d** power-law fit to the data, and are provided as a guide for the eye

discrepancies are observed when using a related technique [3] to determine the most common path between source and target phenotypes (“[Appendix 2: Markov chains to determine the most common path](#)”). Specifically, the predicted most common path matches the empirically observed most common path in only 54 % of the 16×15 combinations of source and target phenotype. In each mismatched case, the length of the observed most common path is greater than or equal to the length of the predicted most common path. For example, when the source phenotype is y and the target phenotype is $!x \text{ AND } y$, the predicted most common path is $y \rightarrow \text{TRUE} \rightarrow !x \text{ AND } y$. In contrast, the empirically observed most common path is $y \rightarrow !y \rightarrow \text{TRUE} \rightarrow !x \text{ AND } y$. This occurs despite the facts that (1) the transition probability encoded in the phenotype network $f_{y,\text{TRUE}} = 0.19$ exceeds that of $f_{y,!y} = 0.185$ and (2) the total probability of the former path exceeds that of the latter.

To elucidate the cause of this discrepancy, we depict in Fig. 12 the average proportion of phenotypically-non-neutral mutations from a genotype in phenotype y to (1) a genotype in phenotype $!y$ and (2) a genotype in phenotype TRUE , as a function of genotypic robustness. Highly robust genotypes are clearly biased toward phenotype $!y$. Since genotypes are visited in proportion to their robustness (Fig. 9b), a random walk through the genotype network of phenotype y is more

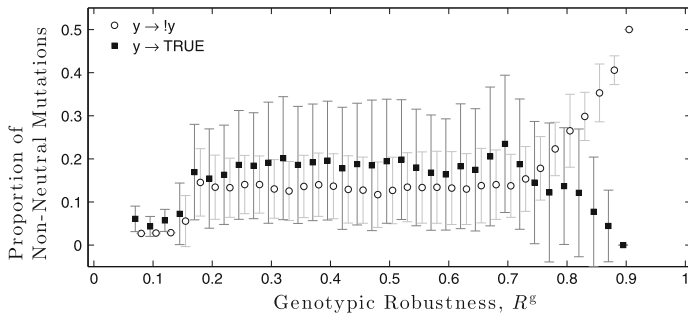


Fig. 12 Phenotypically-non-neutral mutations are not uniformly distributed amongst genotypes. Average proportion of phenotypically-non-neutral mutations from phenotype !y to phenotypes !y (circles) and TRUE (squares), as a function of genotypic robustness R^g . Error bars denote one standard deviation

likely to encounter mutational opportunities to enter phenotype !y than phenotype TRUE despite the fact that in total there are more mutational opportunities to enter the latter than the former. Thus, the transition probabilities encoded in the phenotype network belie the actual mutational biases between genotype networks, and thus violate the assumption that the transition matrix accurately encodes the probability with which one phenotype changes to another.

4 Discussion

Through an exhaustive characterization of genotype networks, this study has described the relationships between robustness, evolvability, and accessibility within and between the genotypic, phenotypic, and fitness levels of a simple LGP system.

At the genotypic level, robustness and evolvability were found to be negatively correlated (Fig. 3c), echoing previous results regarding RNA landscapes [52]. This intuitive observation implies that robust genotypes are located far from the periphery of the genotype network, prohibiting direct mutational access to adjacent genotype networks. Indeed, k -shell decomposition [33] reveals that highly robust genotypes are located in the dense, innermost core of the genotype network (Fig. 3b). This separation of the dense core of highly robust genotypes from the periphery of less robust genotypes (Fig. 3d) also explains the bimodal distribution of genotypic robustness. The peripheral component extends broadly throughout genotype space (Fig. 3e), rendering this representation non-synonymous [40], and the position of a genotype in this space impacts its adjacency to other genotype networks (Fig. 3f).

At the phenotypic level, the distribution of robustness was heterogeneous, indicating that the inherent redundancy of this LGP encoding is non-uniform [40]. Since every phenotype was made up of exactly one genotype network, as opposed to several independent genotype networks, each of the many genotypes that mapped to a given phenotype could be reached via a series of phenotypically-neutral point

mutations. The distribution of phenotypically non-neutral point mutations was also heterogeneous, such that some mutational transitions were more likely than others (Fig. 4). Similar results have been observed in the redundant mappings of alternative evolutionary systems [5, 31, 42].

The mutational transitions between phenotypes varied between the case where deleterious mutations were allowed and the case where they were not allowed. Specifically, when deleterious mutations were allowed, the phenotype network was fully connected, implying that the underlying genotype networks were highly “intertwined,” a feature that is thought to increase evolvability [11, 12]. The exact relationship between robustness and evolvability varied depending on how evolvability was defined (Fig. 5c). When defined by the total connectivity of a phenotype in the phenotype network (E_1^P), evolvability was independent of robustness; all phenotypes were maximally evolvable (cf., Fig. 4). In contrast, when mutational biases were taken into account (E_2^P), the relationship between evolvability and robustness was nonlinear, with phenotypes of intermediate robustness exhibiting the lowest evolvabilities. These results contrast with those made in RNA systems where E_1^P was found to be positively correlated [52], and E_2^P negatively correlated, with robustness [8], providing further evidence that the relationships between these quantities are system-dependent. However, accessibility and robustness were positively correlated (Fig. 5d), in line with observations made in RNA systems and supporting the intuitive notion that phenotypes formed by many genotypes are easier to access than phenotypes formed by few genotypes.

When deleterious mutations were not allowed, as is the case in “replace if better or equal” selection strategies [9, 24], the relationship between phenotypic robustness and phenotypic evolvability was lost, as measured using either E_1^P or E_2^P . Since phenotypes of lower fitness could only mutate into phenotypes of equal or higher fitness, the connectivity of a phenotype in the phenotype network was arbitrarily determined and there was consequently no relationship between these evolvability measures and phenotypic robustness. In contrast, the relationship between phenotypic robustness and phenotypic accessibility remained positive, but with a diminished strength of correlation.

At the fitness level, the distribution of robustness varied between phenotypic targets, as did the mutational connectivities between fitness values (Figs. 6a, b; 8a, b). When the fitness network was constructed under the assumption that deleterious mutations were allowed, there was a weak relationship between the robustness and evolvability of a fitness value (Fig. 7a) due to the distance-based mapping of phenotype to fitness, which arbitrarily grouped phenotypes into fitness values. In contrast, there was a positive correlation between fitness accessibility and fitness robustness (Fig. 7b), again supporting the intuitive notion that what is more common is easier to identify. When the fitness network was constructed under the assumption that deleterious mutations were not allowed, these correlations were completely lost.

As the resolution of analysis shifted from genotype to phenotype, several functional relationships were observed between the two levels. For instance, the means of the distributions of both genotypic robustness and genotypic evolvability were correlated with phenotypic robustness (Fig. 5a, b). Similarly, as the analysis

shifted from phenotype to fitness, the means of the distributions of phenotypic robustness and phenotypic accessibility (Fig. 7b) were correlated with fitness robustness, so long as deleterious mutations were allowed. While some properties were thus correlated between levels, others were not. For example, when deleterious mutations were not allowed, fitness evolvability was independent of robustness at both the phenotypic and fitness levels, a disconnect that again stems from the mapping of phenotype to fitness, in two ways. First, prohibiting deleterious mutations leads to the modification of the phenotype networks so as to only allow mutational transitions from phenotypes of lower fitness to phenotypes of equal or higher fitness (Fig. 8c, d). This arbitrarily severed many of the mutational connections between phenotypes, thus negating any relationship between the robustness and evolvability of a phenotype. Second, the grouping of phenotypes into fitness values without regard to their robustness or evolvability precluded any possibility of a functional relationship between these quantities at the two levels.

To understand how the distributions of robustness, evolvability, and accessibility impact the evolutionary dynamics of LGP, we performed a series of four interrelated analyses. In the first analysis, random walks were used to ascertain the frequency with which a blind evolutionary search would visit a genotype as a function of its robustness. We found that this visit frequency was positively correlated with genotypic robustness (Fig. 9), mirroring classical results for random walks on complex networks [33] and population diffusions on neutral networks [48, 56]. Thus, even in the absence of any selection pressure, blind mutation tends toward increased genotypic robustness.

In the second analysis, we considered an ensemble of random walks between source and target phenotypes. The mean waiting time of random mutation to reach a target phenotype was found to be uncorrelated with the evolvability of the source phenotype (Fig. 10a), a result that calls into question the utility of existing phenotypic evolvability measures. While these measures provide useful information concerning the immediate adjacency of phenotypes [52] and their mutational biases [8], they are too myopic to predict the length of an evolutionary trajectory from one phenotype to another. Consider, for example, that correlations may exist between the evolvabilities of adjacent phenotypes, such that high evolvability phenotypes are mutationally biased toward low evolvability phenotypes. As these correlations (a.k.a. mixing patterns [32]) are not taken into account, the applicability of current phenotypic evolvability measures are left severely constrained, at least for this LGP system. In contrast, the mean waiting time of random mutation to reach a target phenotype was strongly correlated with the target phenotype's accessibility (Fig. 10b). This result provides additional support to earlier suggestions that accessibility is a useful measure for understanding evolutionary dynamics in neutral search spaces [27, 28].

In the third analysis, we assigned fitness values to each phenotype and again considered an ensemble of random walks between source and target phenotypes. The mean waiting time was correlated with both the evolvability of the source fitness value and the accessibility of the target fitness value. That there was a strong correlation between mean waiting time and the evolvability of the source fitness value, but not of the source phenotype, underscores the point that the predictive

power of these measures varies between levels and among evolutionary processes. This point is further supported by the results of our fourth analysis, which was a slight modification of the third. Specifically, the random walk was constrained to always maintain or improve fitness, which corresponds to the case where deleterious mutations are prohibited. In this case, the mean adaptation time was not correlated with the evolvability of the source fitness value (Fig. 11c), but was correlated with the accessibility of the target fitness value (Fig. 11d).

5 Conclusions and future work

There has been much debate regarding the benefit of neutrality in GP. As argued in [16], much of this contention stems from the overly complex problems, representations, and search algorithms used in these investigations, which make it difficult to tease apart the effects of neutrality from other confounding factors. In addition, neutrality is often artificially added to the problem representation and little attention is paid to how this alters the fitness landscape. We circumvented these issues by using a compact representation with inherent neutrality, a simple target-matching problem, and an elementary search algorithm.

Recent experimentation with a minimal GP system has demonstrated that the benefits of neutrality are problem-dependent [15]. Our results provide further support for this hypothesis, as the mean waiting time varied significantly between target phenotypes. For non-uniformly redundant encodings, as considered herein, theoretical models suggest that neutrality only offers an advantage if the optimal phenotype is overrepresented at the genetic level [40]. This prediction is in line with our observation that the mean waiting time decreases as the accessibility of the target phenotype increases, since accessibility is positively correlated with robustness. Our results also extend those of [40] to show that it is not only the uniformity and synonymy of a redundant representation that affect evolutionary search, but also (1) the distributions of robustness at the genotypic and fitness levels and (2) the mutational biases that exist amongst genotypes, phenotypes, and fitness values. Of particular importance is the relationship between the robustness of a genotype and its mutational bias toward other genotype networks.

This study opens the door for several future research directions. First, our analysis can be extended to both larger LGP systems and alternative GP systems. For example, recent investigations in grammatical evolution have demonstrated that the mutational connectivities between phenotypes varies amongst representations and objective functions [31]. Further quantifying these mutational biases using the measures discussed herein could shed additional light on this variability and provide insight into the relationship between robustness, evolvability, and accessibility in another branch of GP. Expanding to these larger and varied systems will require the adoption and modification of the approximation techniques developed for RNA systems [25], as the corresponding genotype networks will not be amenable to exhaustive enumeration. Second, the phenotypic evolvability measures used in this study could be revised to take into account the global structure of a phenotype network, as opposed to only considering the

immediate adjacency of a phenotype. Measures of vertex centrality [33] may provide a useful starting point. Third, the simple evolutionary processes considered in this study can be extended to include population-level processes. This will provide a better understanding of how the structure of genotype, phenotype, and fitness networks impacts the evolutionary dynamics of LGP. Fourth, the role of alternative variation operators, such as recombination, should be considered. Related work on model gene regulatory circuits has demonstrated that recombination can lead to an increase in both robustness and evolvability [29]. However, the sheer breadth of the genotype networks observed in this study implies that this LGP encoding is nonsynonymously redundant, which suggests that recombination may be disruptive in this search space [40]. Nevertheless, understanding how recombination impacts robustness, evolvability, and accessibility in LGP is an important challenge and exciting direction for future work. Lastly, several of the trends revealed in this study may lend themselves to analytical treatment. For example, it may be possible to analytically derive the relationship between the accessibility and robustness of a phenotype $A^P \propto (R^P)^{1/2}$. Earlier analytical results on the convexity of genotype networks [27] may be of relevance in this endeavor. Another relationship that may prove analytically tractable is between the waiting time of a random walk and the accessibility of the target phenotype $T_W \propto (A^P)^{-5/3}$. Such analysis could help generalize our results to other phenotype networks, such as those observed in [49].

In summary, this study has demonstrated that the relationships between robustness, evolvability, and accessibility vary amongst the genotypic, phenotypic, and fitness spaces of LGP, as does the ability of these measures to predict the dynamical properties of an evolutionary process. While mapping the mutational connectivities between phenotypes and fitness values may allow for the development of predictive analytical techniques, such abstractions never tell the complete story; so long as they do not encode the mutational biases that exist at the level of the genotype, their explanatory power will be limited.

Acknowledgments This work was partially supported by NIH grants R01-LM009012, R01-LM010098, and R01-AI59694. J.L.P. was supported by NIH grant R25-CA134286. W.B. acknowledges support from NSERC Discovery Grants, under RGPIN 283304-07. The authors would like to thank the three anonymous reviewers for their scrutiny, Davnah Urbach for her thoughtful comments on earlier versions of this manuscript, and Bill Langdon for fruitful discussions at Evo* 2011.

Appendices

In the appendix, we lay out the details of our analytical treatment, which is based on the techniques described in [3, 17]. Both texts provide exceptionally lucid expositions of absorbing Markov chains and their applications.

Appendix 1: Markov chains to determine the mean first passage time

The mean waiting time and mean adaptation time can be obtained analytically with absorbing Markov chains [17], using the stochastic transition matrix \mathbf{P} whose elements

$$p_{x^p y^p} = \begin{cases} \frac{|\Omega^{x^p, y^p}|}{|\Theta^{x^p}| + \sum_{z \neq y^p} |\Omega^{x^p, z^p}|}, & \text{if } x^p \neq y^p \\ \frac{|\Theta^{x^p}|}{|\Theta^{x^p}| + \sum_{z \neq y^p} |\Omega^{x^p, z^p}|}, & \text{if } x^p = y^p \end{cases} \quad (15)$$

denote the probability of a transition from phenotype x^p to phenotype y^p .

The target phenotype t^p is the absorbing state of the Markov chain; all other phenotypes are transient. The row corresponding to phenotype t^p is therefore modified such that $p_{t^p t^p} = 1$ and $p_{t^p x^p} = 0 \forall x^p \neq t^p$. Placing \mathbf{P} in canonical form, we have

$$\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1t^p} \\ p_{21} & p_{22} & \cdots & p_{2t^p} \\ \vdots & \vdots & \ddots & \vdots \\ p_{t^p 1} & p_{t^p 2} & \cdots & p_{t^p t^p} \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1t^p} \\ p_{21} & p_{22} & \cdots & p_{2t^p} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}, \quad (16)$$

where \mathbf{Q} is a $(|\Phi^p| - 1) \times (|\Phi^p| - 1)$ matrix, \mathbf{R} is a $(|\Phi^p| - 1) \times 1$ column vector, and $\mathbf{0}$ is a $1 \times (|\Phi^p| - 1)$ row vector. To obtain the mean waiting time, we use \mathbf{Q} to calculate the fundamental matrix

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1}, \quad (17)$$

where \mathbf{I} is the identity matrix and entry $n_{x^p y^p}$ is the expected time spent in transient phenotype x^p , given that the random walk started in the transient phenotype y^p . The fundamental matrix is then used to calculate

$$\tau = \mathbf{N}\mathbf{e}, \quad (18)$$

where \mathbf{e} is a column vector of ones and τ_{x^p} is the mean waiting time to reach phenotype t^p from phenotype x^p .

Appendix 2: Markov chains to determine the most common path

Absorbing Markov chains can also be used to determine the most common path from a transient phenotype s^p to a target phenotype t^p . Since the most common path must visit each phenotype at most once, the length of the path must be less than or equal to the total number of phenotypes $|\Phi^p|$. Every path has an associated probability, which can be calculated using the entries of the transition matrix \mathbf{P} . For example, the path $s^p \rightarrow x^p \rightarrow y^p \rightarrow t^p$ has probability $p_{s^p x^p} p_{x^p y^p} p_{y^p t^p}$. The most common path is the one with the highest such joint probability, which can be determined efficiently using the message passing approach described in [3].

References

1. W. Banzhaf, in *Proceedings of Parallel Problem Solving from Nature III*. Genotype-phenotype mapping and neutral variation—a case study in genetic programming. Lecture Notes in Computer Science, vol. 866 (1994) pp. 322–332

2. W. Banzhaf, A. Leier, *Genetic Programming Theory and Practice III, Chapter 14: Evolution on Neutral Networks in Genetic Programming*. (Springer, Berlin, 2006), pp. 207–221
3. D. Barber, *Bayesian Reasoning and Machine Learning*. (Cambridge University Press, Cambridge, 2011)
4. J.D. Bloom, S.T. Labthavikul, C.R. Otey, F.H. Arnold, Protein stability promotes evolvability. *Proc. Natl. Acad. Sci.* **103**, 5869–5874 (2006)
5. S. Bullock, in *Proceedings of the 6th European Conference on Artificial Life*. Smooth operator? Understanding and visualising mutation bias. *Lecture Notes in Computer Science*, vol. 2159 (2001), pp. 602–612
6. S. Ciliberti, O.C. Martin, A. Wagner, Innovation and robustness in complex regulatory gene networks. *Proc. Nat. Acad. Sci.* **104**, 13591–13596 (2007)
7. M. Collins, Finding needles in haystacks is harder with neutrality. *Genet. Prog. Evolvable Mach.* **7**, 131–144 (2006)
8. M.C. Cowperthwaite, E.P. Economo, W.R. Harcombe, E.L. Miller, L.A. Meyers, The ascent of the abundant: How mutational networks constrain evolution. *PLoS Comput. Biol.* **4**(7), e10000110 (2008)
9. B. Dorronsoro, E. Alba, in *IEEE Congress on Evolutionary Computation*. A simple cellular genetic algorithm for continuous optimization (2006), pp. 2838–2844
10. J.A. Draghi, T.L. Parsons, G.P. Wagner, J.B. Plotkin, Mutational robustness can facilitate adaptation. *Nature* **463**, 353–355 (2010)
11. M. Ebner, P. Langguth, J. Albert, M. Shackleton, R. Shipman, in *Proceedings of the IEEE Congress on Evolutionary Computation*. On neutral networks and evolvability (2001), pp 1–8
12. M. Ebner, M. Shackleton, R. Shipman, How neutral networks influence evolvability. *Complexity* **7**(2), 19–33 (2002)
13. E. Ferrada, A. Wagner, Protein robustness promotes evolutionary innovations on large evolutionary time-scales. *Proc. R. Soc. Lond. B* **275**, 1595–1602 (2008)
14. W. Fontana, P. Schuster, Continuity in evolution: On the nature of transitions. *Science* **280**, 1451–1455 (1998)
15. E. Galván-López, R. Poli, in *Proceedings of the Genetic and Evolutionary Computation Conference*. An empirical investigation of how and why neutrality affects evolutionary search (2006), pp. 1149–1156
16. E. Galván-López, R. Poli, A. Kattan, M. O'Neill, A. Brabazon, Neutrality in evolutionary algorithms... What do we know?. *Evol. Syst.* **2**, 145–163 (2011)
17. C. Grinstead, J. Snell, *Introduction to Probability*. (American Mathematical Society, Providence, 1991)
18. I. Harvey, A. Thompson, in *Proceedings of the First International Conference on Evolvable Systems: From Biology to Hardware*. Through the labyrinth evolution finds a way: A silicon ridge (1996), pp. 406–422
19. E.J. Hayden, E. Ferrada, A. Wagner, Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* **474**, 92–95 (2011)
20. T. Hu, W. Banzhaf, in *Proceedings of the Genetic and Evolutionary Computation Conference*. Neutrality and variability: two sides of evolvability in linear genetic programming (2009), pp. 963–970
21. T. Hu, J.L. Payne, W. Banzhaf, J.H. Moore, in *Proceedings of the 14th European Conference on Genetic Programming*. Robustness, evolvability, and accessibility in linear genetic programming. *Lecture Notes in Computer Science*, vol. 6621 (2011), pp. 13–24
22. M.A. Huynen, P.F. Stadler, W. Fontana, Smoothness within ruggedness: the role of neutrality in adaptation. *Proc. Natl. Acad. Sci.* **93**, 397–401 (1996)
23. M. Isalan, C. Lemerle, K. Michalodimitrakis, C. Horn, P. Beltrao, E. Raineri, M. Garriga-Canut, L. Serrano, Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452**, 840–846 (2008)
24. H. Ishibuchi, N. Tsukamoto, Y. Nojima, in *Proceedings of Parallel Problem Solving from Nature X*. Examining the effect of elitism in cellular genetic algorithms using two neighborhood structures. *Lecture Notes in Computer Science*, vol. 5199 (2008), pp. 458–467
25. T. Jörg, O.C. Martin, A. Wagner, Neutral network sizes of biological RNA molecules can be computed and are not atypically small. *BMC Bioinform.* **9**, 464 (2008)
26. M. Kirschner, J. Gerhart, Evolvability. *Proc. Natl. Acad. Sci.* **95**, 8420–8427 (1998)

27. P.K. Lehre, P.C. Haddow, in *Proceedings of the IEEE Congress on Evolutionary Computation*. Accessibility between neutral networks in indirect genotype-phenotype mappings (2005), pp. 419–426
28. P.K. Lehre, P.C. Haddow, in *Simulated Evolution and Learning*. Accessibility and runtime between convex neutral networks. Lecture Notes in Computer Science, vol. 4247 (2006), pp. 734–741
29. O.C. Martin, A. Wagner, Effects of recombination on complex regulatory circuits. *Genetics* **138**, 673–684 (2009)
30. T. Mihaljev, B. Drossel, Evolution of a population of random Boolean networks. *Eur. Phys. J. B* **67**, 259–267 (2009)
31. E. Murphy, M. O'Neill, A. Brabazon, in *Proceedings of the 14th European Conference on Genetic Programming*. Examining mutation landscapes in grammar based genetic programming. Lecture Notes in Computer Science, vol. 6621 (2011), pp. 130–141
32. M. Newman, Mixing patterns in networks. *Phys. Rev. E* **67**, 026126 (2003)
33. M. Newman, *Networks: An Introduction*. (Oxford University Press, Oxford, 2010)
34. M. Newman, R. Engelhardt, Effects of selective neutrality on the evolution of molecular species. *Proc. R. Soc. Lond. B* **265**, 1333–1338 (1998)
35. J.L. Payne, J.H. Moore, in *Proceedings of the European Conference on Artificial Life*. Robustness, evolvability, and accessibility in the signal-integration space of gene regulatory circuits (2011), pp. 638–645
36. D.A. Pechenick, J.L. Payne, J.H. Moore, The influence of assortativity on the robustness of signal-integration logic in gene regulatory networks. *J. Theor. Biol.* **296**, 21–32 (2012)
37. K. Raman, A. Wagner, The evolvability of programmable hardware. *J. R. Soc. Interf.* **8**, 269–281 (2011)
38. C. Reidys, P.F. Stadler, P. Schuster, Generic properties of combinatory maps: Neutral networks of RNA secondary structures. *Bull. Math. Biol.* **59**, 339–397 (1997)
39. J. Rodrigues, A. Wagner, Genotype networks, innovation, and robustness in sulfur metabolism. *BMC Syst. Biol.* **5**, 39 (2011)
40. F. Rothlauf, D.E. Goldberg, Redundant representations in evolutionary computation. *Evol. Comput.* **11**, 381–415 (2003)
41. P. Schuster, W. Fontana, P.F. Stadler, I.L. Hofacker, From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond. B* **255**, 279–284 (1994)
42. R. Shipman, M. Shackleton, M. Ebner, R. Watson, in *Proceedings of the 7th International Conference on Artificial Life*. Neutral search spaces for artificial evolution: a lesson from life (2001), pp. 162–169
43. T. Smith, P. Husbands, P. Layzell, M. O'Shea, Fitness landscapes and evolvability. *Evol. Comput.* **10**, 1–34 (2002)
44. T. Smith, P. Husbands, M. O'Shea, in *Proceedings of the 6th European Conference on Artificial Life*. Neutral networks and evolvability with complex genotype-phenotype mapping. Lecture Notes in Computer Science, vol. 2159 (2001), pp. 272–282
45. T. Smith, P. Husbands, M. O'Shea, in *Proceedings of the IEEE Congress on Evolutionary Computation*. Neutral networks in an evolutionary robotics search space (2001), pp. 136–145
46. T. Soule, Resilient individuals improve evolutionary search. *Artif. Life* **12**, 17–34 (2006)
47. B.M.R. Stadler, P.F. Stadler, G.P. Wagner, W. Fontana, The topology of the possible: Formal spaces underlying patterns of evolutionary change. *J. Theor. Biol.* **213**, 241–274 (2001)
48. van E. Nimwegen, J. Crutchfield, M. Huynen, Neutral evolution of mutational robustness. *Proc. Natl. Acad. Sci.* **96**, 9716–9720 (1999)
49. L. Vanneschi, Y. Pirola, P. Collard, M. Tomassini, S. Verel, G. Mauri, in *Proceedings of the Genetic and Evolutionary Computation Conference*. A quantitative study of neutrality in GP Boolean landscapes (2006), pp. 895–902
50. A. Wagner, *Robustness and Evolvability in Living Systems*. (Princeton University Press, Princeton, 2005)
51. A. Wagner, Neutralism and selectionism: a network-based reconciliation. *Nat. Rev. Genet.* **9**, 965–974 (2008)
52. A. Wagner, Robustness and evolvability: a paradox resolved. *Proc. R. Soc. Lond. B* **275**, 91–100 (2008)
53. A. Wagner, *The Origins of Evolutionary Innovations*. (Oxford University Press, Oxford, 2011)
54. G. Wagner, L. Altenberg, Complex adaptation and the evolution of evolvability. *Evolution* **50**, 967–976 (1996)

55. J. Whitacre, A. Bender, Degeneracy: a design principle for achieving robustness and evolvability. *J. Theor. Biol.* **263**, 143–153 (2010)
56. C.O. Wilke, Adaptive evolution on neutral networks. *Bull. Math. Biol.* **63**, 715–730 (2001)
57. T. Yu, J.F. Miller, Through the interaction of neutral and adaptive mutations, evolutionary search finds a way. *Artif. Life* **12**, 525–551 (2006)