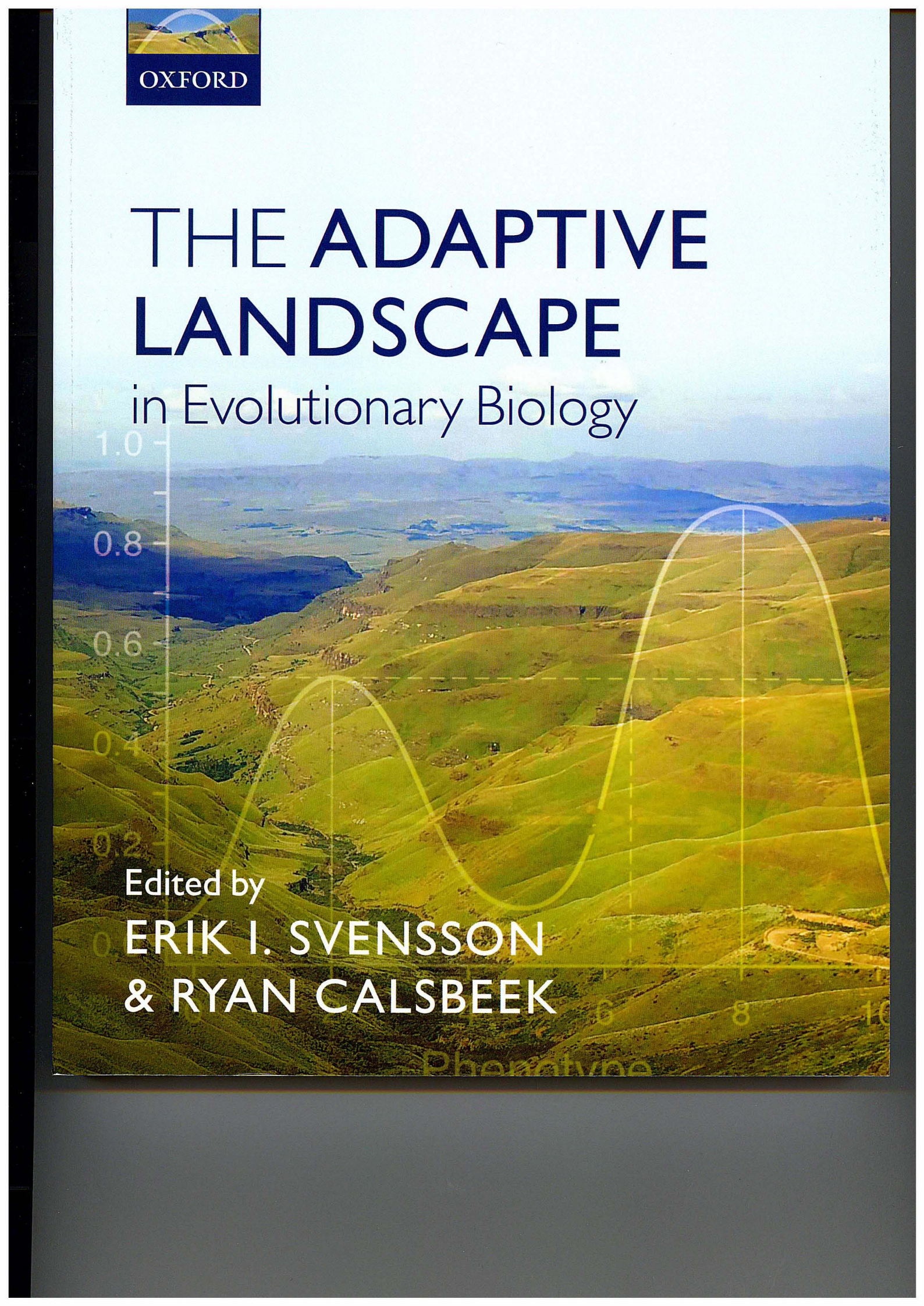
OXFORD
THE ADAPTIVE LANDSCAPE
in Evolutionary Biology
1.0
0.8
0.6
0.4
0.2
0
Edited by
ERIK I. SVENSSON
& RYAN CALSBEEK

CHAPTER 17

High-Dimensional Adaptive Landscapes Facilitate Evolutionary Innovation

Andreas Wagner

17.1 Introduction

The Adaptive Landscape is one of the most influential concepts in evolutionary biology (Wright 1932, Futuyma 1998, see also Chapter 2 of this volume). It is commonly visualized as a surface of rolling hills or rugged mountains in a three-dimensional space. Two of the three dimensions represent allele frequencies or—relevant for my purpose—genotypes. The third dimension represents fitness. The landscape's peaks represent adaptive trait combinations or genotypes. The Adaptive Landscape concept has been highly successful, so much so that it has spawned multiple variants in the hands of different authors, including holey landscapes (Gavrilets 1997) and phenotype landscapes. The variant I emphasize here can be viewed as a phenotype landscape (see also Chapter 18 of this volume for different kinds of phenotype landscapes).

I view the Adaptive Landscape as a metaphor for the evolutionary process. It is an abstraction derived from an immensely complex reality. Such abstraction is necessary for all human understanding of the world around us. Yet like all abstractions, it also has limitations. Several of these limitations are caused by the high dimensionality of genotype spaces. This high dimensionality has already been appreciated by Sewall Wright, the creator of the Adaptive Landscape concept (Wright 1932). Its consequences have been studied at least since the late 1980s (Kauffman and Levin 1987; Conrad 1990; Gavrilets 1997). One important such consequence regards a basic geometric intuition we derive from three-dimensional space. To get from one peak to the next, one needs to cross a maladaptive valley—the valley of death, if you will—with no detour around this valley. This changes in higher dimensional landscapes, where, counterintuitively, "extradimensional bypasses" around maladaptive valleys exist (Conrad 1990). Through such bypasses, one can travel from peak to peak while avoiding valleys of death. Gavrilets pointed out that this principle has implications for the evolution of reproductive isolation (Gavrilets 1997, 2005). I here explain that it also has implications for how biological systems innovate.

Innovations in evolving biological systems are qualitatively new and beneficial new phenotypes. High-dimensional Adaptive Landscapes can facilitate such innovations. In the next section I discuss evidence for this assertion for three classes of systems important for evolutionary innovation. These are metabolic networks, gene regulation circuits, as well as protein and RNA macromolecules. Detailed studies of the high-dimensional genotype spaces of these systems have demonstrated the existence of vast genotype networks. These are connected sets of genotypes with the same phenotype that extend far through genotype space. Genotype networks are important for the ability of biological systems to explore many novel phenotypes. Genotype networks can be viewed as a consequence of the high-dimensionality of genotype spaces. I will begin illustrating this feature in some detail with metabolic networks, and then discuss the other system classes more briefly. Table 17.1 summarizes some of the concepts I will introduce.

Table 17.1 Important concepts for the three system classes discussed in this chapter.

	Genotype	Neighbors	Genotype space	Phenotype
Metabolic Network	DNA encoding all metabolic enzymes/reactions in a metabolism	Networks that differ in one enzyme (reaction)	All possible metabolic networks	Ability to synthesize biomass or other important molecules from a given spectrum of nutrients
Regulatory circuit	DNA encoding regulatory interactions among molecules	Circuits that differ in one regulatory interaction	All possible regulatory circuits	Gene expression or molecular activity of all circuit molecules
Macromolecule (protein/RNA)	Amino acid or nucleotide polymers	Polymers that differ in a single monomer	All possible amino acid/nucleotide polymers	Tertiary structure (fold) and biochemical activity

17.2 Metabolic network space

Metabolic networks are comprised of hundreds to thousands of chemical reactions—catalyzed by enzymes that are encoded by genes—which synthesize all small molecules in biomass from environmental nutrients. In addition, they produce energy and many important secondary metabolites. Such networks are involved in many innovations, from microbes to higher organisms. Examples include the ability of microbes to grow on synthetic antibiotics or other toxic xenobiotic compounds, such as polychlorinated biphenyls, chlorobenzenes, or pentachlorophenol, in some cases using them as their sole carbon source (Cline et al. 1989; van der Meer 1995; van der Meer et al. 1998; Copley 2000; Dantas et al. 2008; Rehmann and Daugulis 2008). They also include the urea cycle, a metabolic innovation of land-living animals that allows them to convert toxic ammonia into urea for excretion. Novel metabolic traits often involve new combinations of chemical reactions (enzymes) in an organism that already exist separately elsewhere. For example, a novel metabolic pathway to degrade pentachlorophenol involves four steps that its host organism assembled—probably through horizontal gene transfer—from enzymes processing naturally occurring chlorinated chemicals, as well as from an enzyme involved in tyrosine metabolism (Copley 2000). Similarly, the urea cycle arose when four widespread enzymatic reactions involved in arginine biosynthesis combined with arginase, a reaction involved in arginine degradation (Takiguchi et al. 1989).

To study innovation in metabolic networks systematically, one needs to be able to represent all possible metabolic networks and their biosynthetic abilities. To do so, it is necessary to define a space of possible metabolic genotypes. The metabolic genotype of any one organism is the part of the organism's genome that encodes the enzymes which catalyze metabolic reactions. Although this genotype is a string of DNA, it is useful to represent it in a more compact way as follows. Consider the known "universe" of enzyme-catalyzed chemical reactions. This universe currently comprises more than 5000 reactions and the associated enzymes. Write the names of these reactions or their stoichiometric equations as a list. For any one organism, write a one next to the reaction in this list if its genome encodes an enzyme catalyzing this reaction, and a zero if it does not. The result will be a long string of ones and zeroes that can represent the metabolic genotype of the organism (Rodrigues and Wagner 2009).

In this representation, a metabolic genotype can be viewed as a single point in a vast metabolic *genotype space*, a space of possible metabolic networks. This space contains all possible presence/absence combinations of enzyme-catalyzed reactions. Each such combination constitutes a metabolic network. For a universe of 5000 reactions, there are 2^{5000} such metabolic networks, a hyperastronomical number, many more than could be realized in organisms on earth. Metabolic genotype space can also be viewed as the set of all binary strings of length 5000. Yet another, geometric representation is that of a 5000-dimensional hypercube graph. This is a graph whose vertices—the vertices of an n-dimensional cube—are the individual genotypes. Two genotypes are connected by an edge if

they are (1-mutant) neighbors, that is, if they differ in a single chemical reaction.

This genotype space is a prototypical example of a high-dimensional space. Each of its axes (reactions) corresponds to one dimension. In contrast to our low-dimensional Euclidean space, it is a discrete and not a continuous space, containing an enumerable number of elements. It is thus quite different from Euclidean space. However, it also shares some similarities with this space. The most important of them is that in genotype space intuitive measures of the *distance* between two metabolic genotypes exists. One such measure is simply the fraction of metabolic reactions in which two genotypes differ. In mathematical terms, metabolic genotype space is thus a *metric* space (Searcoid 2007).

17.3 From metabolic genotype to phenotype

A free-living organism such as *Escherichia coli* or yeast needs to synthesize some 50 essential biomass molecules to grow and divide (Forster et al. 2003; Feist et al. 2007). These molecules include all 20 proteinaceous amino acids, RNA and DNA nucleotides, lipids, and enzyme co-factors.

To sustain life, the metabolic network of a heterotrophic organism needs to generate energy and synthesize all these molecules from a limited number of chemicals in the environment. Recent advances in computational methods have made it possible to compute whether a metabolic network can do so, including the rate at which it can synthesize each compound (Schilling et al. 1999; Feist and Palsson 2008; Feist et al. 2009). For this computation, one needs two kinds of information. These are the stoichiometric equations for each of the chemical reactions in the network, and the rate at which an organism can import necessary chemicals from the environment. Given this approach, one can compute metabolic phenotypes from metabolic genotypes.

To study metabolic innovation is to study how qualitatively novel metabolic phenotypes arise. This requires a definition of phenotype that is suitable for a systematic analysis, and suitable to compare phenotypes. There are many ways to define a metabolic phenotype. For example, one could list the essential biomass molecules that a metabolic network is able to synthesize in any one chemical environment. However, unless a metabolic network can synthesize *all* essential molecules, it may not be able to sustain life. This definition is therefore of limited use. To study metabolic innovations, and especially innovations that allow survival on novel sources of energy and carbon, the following definition is more useful. This definition focuses on carbon, because carbon is a central element in life, and because innovations involving carbon metabolism are thus especially important. However, what I describe also applies to other metabolic innovations (Rodrigues and Wagner 2011).

Consider a minimal chemical environment that contains a small number of molecules which can serve as a source of all necessary elements except carbon. For instance, in the case of *E. coli*, this environment comprises only six different kinds of molecules. Now make a list of many different potential sources of carbon and energy, such as glucose, ethanol, glycerol, and so forth. For the sake of the argument, let us consider 100 such sources. For each of these sources, when provided in an otherwise minimal environment as the *sole* carbon source, determine whether a given metabolic network can synthesize all essential biomass metabolites. If so, write a one next to the list of carbon sources. If not, write a zero. Define the resulting string as the *metabolic phenotype* of this metabolic network. It represents the set of carbon sources on which the metabolic network can sustain life, on which it is *viable* (Rodrigues and Wagner 2009).

This definition of a metabolic phenotype is well-suited for a systematic analysis and comparison of phenotypes, including metabolic innovations. First, it encapsulates an astronomical number of different phenotypes (2^{100} for 100 carbon sources). Second, this notion of phenotype makes it easy to compare different phenotypes by comparing their associated binary strings. Third, an evolutionary innovation—viability on new carbon sources—simply corresponds to a phenotype string where one or more zeroes are converted to ones. It fits the definition of an innovation as a new phenotype that can make a qualitative difference to survival in the right environments, namely environments where

this carbon source is the only available source. Such a novel phenotype can arise by adding reactions to a metabolic network, for example by adding enzyme-coding genes to a genome via horizontal gene transfer.

17.4 Exploring metabolic genotype space

To characterize metabolic genotype space exhaustively is impossible, but one can obtain much insight into the organization of this space by carefully designed random sampling. One relevant class of approach is Markov chain Monte Carlo sampling, which involves random walks through genotype space (Rodrigues and Wagner 2009; Samal et al. 2010). Such random walks modify a starting network in a series of steps, each of which either eliminates a reaction (such as might occur through a loss-of-function mutation in an enzyme coding gene) or adds a reaction (such as might occur through horizontal gene transfer). It is useful to require that each step of such a random walk preserves the phenotype. During this random walk, one can also determine all metabolic genotypes in the immediate neighborhood of the random walking network, determine their metabolic phenotypes, and ask which of them are novel metabolic phenotypes that allow survival on new carbon sources. This neighborhood is of special interest, because it contains all novel metabolic phenotypes that are easily accessible from a network through changes in a single reaction.

We have carried out such random walks from different starting points, with very different starting metabolic phenotypes, and explored innovations in the utilization of carbon and other elements (Rodrigues and Wagner 2009; Samal et al. 2010). Together, these analyses have revealed some strikingly simple principles of metabolic genotype space organization.

First, individual metabolic genotypes typically have many neighbors with the same metabolic phenotype. In other words, the metabolic phenotypes of these metabolic networks are to some extent *robust* to mutations that involve changes in individual reactions.

Second, genotypes with the same phenotype form vast connected genotype networks that reach far through genotype space. This means that one can step from one genotype to its neighbor, to the neighbor's neighbor, and so on, without ever changing a phenotype. A genotype network can be viewed as a network of metabolic networks in genotype space. Two genotypes that are far apart on this network have the same phenotype but may share fewer than 25% of their chemical reactions (Rodrigues and Wagner 2009).

Third, the neighborhood of any two genotypes on the same genotype network contains very different novel phenotypes (Fig. 17.1). Together, these properties facilitate the evolution of novel phenotypes through exploration of genotype space. They allow a population to keep its phenotype unchanged while exploring different regions of genotype space and many novel phenotypes therein.

Genotype networks are not peculiarities of carbon metabolism. They also occur in the metabolism of other elements (Rodrigues and Wagner 2011). I note in passing that they have very similar properties if one requires that each random walk preserves the rate at which a network can synthesize biomass, and not just the mere ability to synthesize biomass (Samal et al. 2010).

17.5 Regulatory circuits and novel gene expression patterns

The second system class I will briefly discuss here are regulatory circuits. They are systems of interacting gene products that influence each other's biological activity. Their phenotypes are gene expression phenotypes or, more generally, molecular activities of gene products with important biological functions. Such circuits are involved whenever cells and tissues communicate, and whenever gene expression is regulated (Gilbert 1997; Carroll et al. 2001). Both processes are indispensable for the development of any multicellular organism, and thus for the formation of all macroscopic phenotypes. The most important kinds of circuits are transcriptional regulation circuits, because transcriptional regulation provides a regulatory backbone of all organismal life, and because such circuits drive many pattern formation processes in

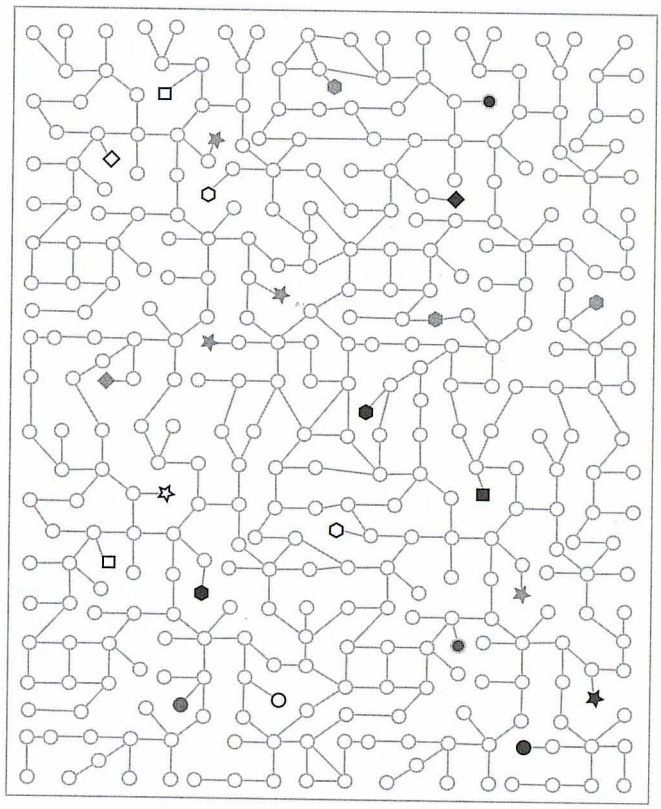

Figure 17.1 Genotype networks in genotype space. The large rectangle schematically represents a genotype space. Open circles represent genotypes with the same (hypothetical) phenotype; neighboring genotypes are connected by straight lines. The figure shows that the open circles form a large connected network that extends far through genotype space. The figure also contains symbols of various shapes and shading. Each such symbol represents a genotype with a different new phenotype, where this genotype is adjacent to the genotype network. That is, it can be reached through a single, small genetic change from some genotype on this network. The figure illustrates that many different novel phenotypes can be accessed from a connected genotype network that spreads far through genotype space. The usual caveat that two-dimensional images poorly represent high-dimensional spaces applies. For example, each genotype typically has hundreds to thousands of neighbors, many more than can be shown here. Also, each genotype that is not on the focal genotype network (symbols of different shape and shading) is also part of a large genotype network that is not shown here. Figure from Wagner (2011), used with permission from Oxford University Press.

embryonic development. Among the best known examples are Hox gene circuits involved in patterning limbs and many other body structures in animals, as well as circuits involving MADS-box genes important in patterning flowers (Hughes and Kaufman 2002; Irish 2003; Wagner et al. 2003; Causier et al. 2005; Lemons and McGinnis 2006; Hueber and Lohmann 2008).

Regulatory change in transcriptional regulation circuits and other circuits is involved in forming many new novel macroscopic phenotypes. For example, the formation of dissected leaves, an innovation of some plants that may aid thermoregulation, is driven by overexpression of KNOX (KNOTTED1-like homeobox) transcription factors in developing leaves (Bharathan et al. 2002). The predator-deterring eye-spots of butterflies form where the transcription factor *Distal-less* is overexpressed. In these and many other examples (Carroll et al. 2001), changes in the regulation of existing molecules (that also serve other, ancestral functions) are associated with the formation of novel phenotypes.

Because transcriptional regulation circuits are central to embryonic development and to regulatory innovations, many such circuits are well-studied individually (Carroll et al. 2001; Davidson and Erwin 2006). However, a systematic understanding of novel regulatory phenotypes requires analysis of not one circuit, but a systematic analysis of thousands of circuits in the *genotype space* of such circuits. For this purpose, computational models of such circuits are currently indispensable (von Dassow et al. 2000; Albert and Othmer 2003; MacCarthy et al. 2003; Jaeger et al. 2004; Sanchez et al. 2008).

Models that lend themselves to an exploration of a circuit space represent the topology of such a circuit, that is, the pattern of activating and inhibiting regulatory interactions, in a systematic way (MacCarthy et al. 2003; Wagner 2005a; Ciliberti et al. 2007a, 2007b). One can think of a circuit's pattern

of regulatory interactions—mediated by the DNA that encodes transcription factors and *cis*-regulatory regions—as the *regulatory genotype* of such a circuit (Fig. 17.1). For a transcriptional regulation circuit, this genotype describes which genes activate or repress each other's expression, and how strongly they do so. Each circuit genotype exists in a space of possible such genotypes. This space contains circuits with all possible patterns of regulatory interactions between circuit genes. Two circuits are neighbors in this space if they differ in exactly one regulatory interaction. The distance between two circuits is the number or fraction of regulatory interactions in which they differ. Two circuits are maximally different, if they have no regulatory interactions in common. The regulatory interactions specified in a regulatory genotype determine the circuit's *phenotype*. This phenotype reflects the activity or expression level of each gene in any one cell, which can be represented either continuously or discretely ("on" or "off"). This highly simplified discrete representation can facilitate enumeration and comparison of different circuit phenotypes (Ciliberti et al. 2007a, 2007b).

Just as in the case of metabolic systems, exploration of the genotype space of circuits shows features very similar to metabolic network space (MacCarthy et al. 2003; Wagner 2005a; Ciliberti et al. 2007a, 2007b; Giurumescu et al. 2009). First, circuits typically have many neighbors in circuit space with the same expression phenotype. Second, circuits with the same phenotype form vast connected genotype networks (MacCarthy et al. 2003; Wagner 2005a; Ciliberti et al. 2007a). Third, the neighborhoods of different circuits contain very different novel phenotypes (Ciliberti et al. 2007b; Giurumescu et al. 2009).

17.6 Macromolecules

The final system class are macromolecules—protein and RNA. They form enzymes, exchange chemicals between cells and their environment, give structural support to cells, are central to locomotion and transport, and perform many other essential functions. Not surprisingly then, many adaptive phenotypic changes are directly traceable to changes in macromolecules. One example regards the ability of some animals to survive temperatures where normal body fluids would freeze. This ability is caused by antifreeze proteins, evolutionary innovations that arose rapidly, multiple times independently, and from different ancestors in arctic and Antarctic fish (Chen et al. 1997; Cheng 1998). Another example involves the ability of the bar-headed goose (*Anser indicus*) to migrate over the Himalayas at altitudes exceeding 10 kilometers (Golding and Dean 1998; Liang et al. 2001). In this bird, one of the subunits of hemoglobin experienced a single proline to alanine substitution. This substitution increases the protein's affinity to oxygen, and thus allows it to transport oxygen at lower oxygen concentrations. It is one of several changes that make the bar-headed goose one of the highest flying birds.

The genotype space of macromolecules is the space of all possible nucleotide or amino acid sequences (Table 17.1). Its structure has been studied for many years (Lipman and Wilbur 1991; Schuster et al. 1994), and shows the same three features I discussed earlier for other genotype spaces. First, individual genotypes typically have many neighbors with the same phenotype (Reidys et al. 1997; Sumedha et al. 2007). For example, random mutagenesis studies of different proteins showed that a large fraction of amino acid changes do not affect protein function (Kleina and Miller 1990; Rennell et al. 1991; Huang et al. 1996).

Second, genotype networks exist in these spaces. This has first been demonstrated for lattice protein models of protein folding, and later for real proteins (Babajide et al. 1997; Todd et al. 1999, 2001; Bastolla et al. 2003; Wagner 2005b). It has also been shown for secondary structure phenotypes of RNA, where genotype networks have been called neutral networks and are extensively characterized (Schuster et al. 1994; Schuster 2006). (I note parenthetically that there are good reason not to use the term neutral network in this context, because evolution along a genotype network need not be neutral in the molecular evolutionist's sense (Wagner 2008).)

Genotype networks in macromolecules typically also extend far through genotype space. The differences between proteins with the same tertiary structure phenotype and common ancestry, for example, can be dramatic. Such proteins may share only a few per cent of their amino acids (Goodman et al.

1988; Rost 1997; Thornton et al. 1999; Todd et al. 1999; Copley and Bork 2000, Bastolla et al. 2003).

Third, different neighborhoods of a genotype network harbor different novel phenotypes. This has first been shown for secondary structure phenotypes of RNA and more recently for proteins and their enzymatic function phenotypes (Schuster et al. 1994; Huynen et al. 1996; Sumedha et al. 2007; Ferrada and Wagner 2010).

17.7 Genotype networks as a consequence of high dimensionality

In sum, three very different classes of biological systems, all of them central for evolutionary innovation, show very similar organizations of their genotype spaces. First, in all three system classes, genotypes have many neighbors with the same phenotype. Specifically, between 10% and more than 50% of a genotype's neighbors typically have the same phenotype, depending on system class, system size, genotype, and phenotype (Wagner 2011). In other words, these systems are to some extent robust to genetic change. Second, genotypes with the same phenotype form vast genotype networks that reach far through genotype space. They typically span between 70–100% of the diameter of this space (Wagner 2011). I note that even though genotype networks are usually astronomically large (10^{50} genotypes in a genotype network are not unusual), and even though they reach far through genotypes space, any one genotype network typically occupies only a vanishing fraction of genotype space. That these properties do not contradict each other results from the fact that genotype space has many dimensions and that it is vast—it has room for myriad genotype networks that are tightly interwoven (Wagner 2011). Third, different neighborhoods of a genotype network generally contain different novel phenotypes. Fig; 17.1 shows a schematic of one such network (open circles) and some novel phenotypes in its neighborhood (symbols of various shapes).

The second and third feature *jointly* facilitate the exploration of novel phenotypes. Specifically, they solve a major problem that innovation poses to living systems: organisms need to preserve old, adaptive phenotypes while exploring innumerable new phenotypes, only few of which may be improvements over the old. Envision a population of organisms that preserves its existing phenotype (through stabilizing selection) while being exposed to mutational change. The existence of genotype networks means that the population can gradually change its genotype while preserving its phenotype. In doing so, it can explore different regions of genotype space. The immediate neighborhood of the population will contain very different novel phenotypes, depending on where its members are located in genotype space. The existence of genotype networks, combined with the diversity of their neighborhoods thus allows exploration of a myriad novel phenotypes.

I note that the discretization of genotypes and phenotypes I used here serves to develop necessary concepts clearly. It also facilitates computational analysis of complex phenotypes and their organization in genotype space. However, a small but growing body of research hints that these concepts can be transferred to systems with continuous genotypes and phenotypes to study how new phenotypes arise in such systems (Wagner 2005a; Giurumescu et al. 2009; Hafner et al. 2009; Raman and Wagner 2011).

The genotype–phenotype relationships I discussed here can be viewed as functions from high-dimensional genotype spaces to high-dimensional spaces of phenotypes. In the case of metabolic networks, the genotypes reflect the presence or absence of metabolic reactions from a reaction universe in any one metabolic network, and the phenotypes reflect the set of carbon sources (or sources of other elements) on which the metabolic network can sustain life. For regulatory circuits, genotypes are the DNA sequences that encode a circuit's regulatory interactions, and phenotypes are activity or concentration patterns of molecules. For macromolecules, genotypes are amino acids and nucleotide sequences, and phenotypes correspond to complex three-dimensional folds of these molecules and their biochemical function (Table 17.1).

I also note that for the systems I consider here, there will generally be more genotypes than phenotypes. For example, for proteins that are N amino acids long, there is an astronomical number of 20^N

genotypes even for moderately large N. In contrast, the number of protein tertiary *structure* phenotypes (protein folds) is of the order 10^4 (Levitt 2009), and for enzymes, the most prominent class of proteins, the number of known *function* phenotypes—the number biochemical reactions they catalyze—is of the same order of magnitude (Ogata et al. 1999). Even if these estimates were to be too low by a factor 100 or 1000, the total number of protein phenotypes would be minute compared to the number of genotypes. In regulatory circuits and metabolism, different arguments lead to the same conclusion (Wagner 2011). For example, in regulatory circuits involving N molecules, the number of regulatory genotypes scales with the number of possible interactions between molecules, and thus with N^2, whereas the number of possible activity states scales with the number of molecules N. There will therefore be more regulatory genotypes than phenotypes. I finally note that if the number of genotypes did not exceed the number of phenotypes, be they protein structures, gene expression phenotypes, metabolic phenotypes, or visible macroscopic traits of organisms, then neutral mutations would not exist, contrary to what empirical genetic data suggest (Eyre-Walker and Keightley 2007).

Any systematic analysis of innovation requires phenotypes that are complex, multidimensional objects, and not just simple scalars, as in many population genetic models of evolution. The phenotypes in all three systems I discuss here meet this criterion. For example, metabolic phenotypes can be represented as vectors in many dimensions, as can the spatial coordinates of a folded protein's amino acids.

The concepts I discuss here share one commonality with holey Adaptive Landscapes used to study speciation (Gavrilets 1997), but they differ in more important ways. The commonality is that genetic change can occur along ridgelines in a high-dimensional space.

The first major difference regards the multidimensional nature of the phenotypes I consider. To study speciation and reproductive isolation, it is may be adequate to consider scalar-valued fitness—or, as in holey landscapes, binary fitness of values 0 and 1—as the only aspect of phenotype. (In this case, the genotypes with fitness zero correspond to the holes in the landscape.) But to study innovation, this will not suffice. It becomes necessary to study phenotypes as complex, multidimensional objects. One could say that for innovation, it is the off-ridge regions in the landscape that are most important (although they are no longer mere holes). They represent genotypes with new phenotypes, some of which may have superior fitness. The second major difference is that to understand innovation, a detailed mechanistic understanding of how phenotypes emerge from genotypes is crucial. The system classes I discuss here meet this requirement. It is not needed to study speciation on holey landscapes, where phenotypes (fitness) are typically assigned *randomly* to genotypes. Finally, I note that innovations also occur in organisms where reproductive isolation, and thus also reproductive isolation by the holey-landscape mechanism, has limited relevance. They include asexual eukaryotes and prokaryotic microbes with unusual forms of sex. Innovation does not generally require reproductive isolation.

How does a map from complex multidimensional genotypes to a multidimensional phenotypes relate to an Adaptive Landscape? To see the connection, we need to simplify this map a bit. Consider the example of metabolic systems, and a specific metabolic phenotype P. For this phenotype, one can define an Adaptive Landscape in metabolic genotype space as a function from this space into the non-negative real numbers. Each genotype's value of this function (its height in the landscape) maps onto the distance of its phenotype from P. Peaks of the landscape correspond to genotypes whose phenotypes are equal to P. Genotypes whose phenotype have a given distance from P occur along the same contour lines (elevation) of the landscape.

Analogous definitions are possible for genotype–phenotype maps in regulatory systems and in macromolecules, because one can define analogous distances among their phenotypes (e.g. Schuster et al. 1994; Ciliberti et al. 2007a). Exactly as for metabolic systems, the height of a point (genotype) corresponds to the distance its phenotype has from some focal phenotype. The genotypes with the focal phenotype are the peaks in this landscape. The more distant a genotype's phenotype is from this focal

phenotype, the lower its elevation in this landscape. This simplification renders the maps I consider instances of phenotype landscapes with a scalar phenotype (see also Chapter 18 of this volume).

In Adaptive Landscapes thus defined, the existence of genotype networks clearly violates our geometric intuition derived from low-dimensional spaces. For any one phenotype, a genotype network corresponds to a series of peaks that occur throughout genotype space, and that are all connected to one another. No valleys separate these peaks. In other words, they form an interlaced network of ridges reaching into distant corners of the space. To confound our intuition further, this web of ridges exists for many different focal phenotypes. Each of them has its own genotype network, and all these genotype networks are tightly interwoven with one another (Schuster et al. 1994; Ciliberti et al. 2007b; Rodrigues and Wagner 2009).

Fundamentally, the reason why genotype networks exist is that typical genotypes with some phenotype P have many neighbors with the same phenotype. If this was not the case, and if the genotypes comprising a typical genotype network were otherwise randomly distributed in genotype space, these genotypes would be isolated from one another (Ciliberti et al. 2007a; Wagner 2011). In other words, their robustness to mutation brings forth the vast genotype networks of which they are a part.

The fact that a genotype can have many neighbors at all emerges from the high-dimensionality of genotype space. In a reaction universe of 5000 reactions, each metabolic network has 5000 neighbors. In a transcriptional regulation circuit of 20 genes, there are of the order of 400 possible regulatory interactions; each circuit thus has of the order of 400 neighbors. In a protein of 100 amino acids and 20 possible amino acids, each genotype has $100 \times 19 = 1900$ numbers of neighbors. Because large numbers of neighbors are possible only in a high-dimensional space, so is the possibility to have many neighbors with the same phenotype, and thus the existence of vast genotype networks.

17.8 Outlook

Applying the Adaptive Landscape concept to low-dimensional genotype spaces and to simple, scalar phenotypes limit its utility. This does not mean that we should abandon the concept. There may be philosophical quibbles about it (Provine 1986, discussed also in Chapter 2), but nothing speaks better to its success than its widespread usage almost a century after its conception.

The various incarnations the Adaptive Landscape has taken in the hands of different researchers (e.g., Chapters 2, 3, 9, 13, and 19) show that rather than abandoning the concept, we need to refine it for specific purposes. To study innovation, for example, we can extend it to functions on genotypes whose values are qualitatively different phenotypes in a high-dimensional space. In that case, mathematical or computational analysis needs to replace our geometric intuition. Further refinements will undoubtedly be necessary. For example, some phenotypes and even genotypes are best viewed as objects in a continuous and not in a discrete space. Examples do not only include models from quantitative genetics where genotypes are continuously valued genetic "variables" that underly a uni- or multivariate continuous phenotype. They also include, on the genotypic level, the regulatory genotypes of many signaling circuits, which are defined through parameters such as (continuous) association and dissociation constants, and reaction rates.

On the phenotypic level, they include the ever-changing conformations—defined by the continuous atomic coordinates of amino acids—that one protein can adopt through thermal noise, or the effectively continuous changes of concentrations that gene products can undergo. Where such continuous systems have been studied, phenomena analogous to those in discrete systems seem to exist (Wagner 2005a; Giurumescu et al. 2009; Hafner et al. 2009). However, we lack the theoretical foundation to study the concepts I emphasized here rigorously in such systems. For example, even just defining the analogue to a genotype network in a continuous, high-dimensional genotype space presents challenges. If we meet these challenges, we may discover new worlds in the vast universe of genotype space. There is little doubt that Adaptive Landscapes, despite their limitations, will help us in understanding this universe.

Acknowledgments

I thank Charles Goodnight, the editors, and an anonymous reviewer for suggestions that helped me improve this manuscript. I also acknowledge support through Swiss National Science Foundation grants 315200–116814, 315200–119697, and 315230–129708, as well as through the YeastX project of SystemsX.ch, and the University Priority Research Program in Systems Biology at the University of Zurich.

References

Albert, R. and Othmer, H.G. (2003). The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. Journal of Theoretical Biology, 223, 1–18.

Ancel, L.W. and Fontana, W. (2000). Plasticity, evolvability, and modularity in RNA. Journal of Experimental Zoology/Molecular Development and Evolution, 288, 242–283.

Babajide, A., Hofacker, I., Sippl, M., and Stadler, P. (1997). Neutral networks in protein space: a computational study based on knowledge-based potentials of mean force. Folding & Design, 2, 261–269.

Bastolla, U., Porto, M., Roman, H.E., and Vendruscolo, M. (2003). Connectivity of neutral networks, overdispersion, and structural conservation in protein evolution. Journal of Molecular Evolution, 56, 243–254.

Bharathan, G., Goliber, T.E., Moore, C., Kessler, S., Pham, T., and Sinha, N.R. (2002). Homologies in leaf form inferred from KNOXI gene expression during development. Science, 296, 1858–1860.

Carroll, S.B., Grenier, J.K., and Weatherbee, S.D. (2001). From DNA to diversity. Molecular genetics and the evolution of animal design. Blackwell, Malden, MA.

Causier, B., Castillo, R., Zhou, J.L., Ingram, R., Xue, Y., Schwarz-Sommer, Z., et al. (2005). Evolution in action: Following function in duplicated floral homeotic genes. Current Biology, 15, 1508–1512.

Chen, L.B., DeVries, A.L., and Cheng, C.H.C. (1997). Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. Proceedings of the National Academy of Sciences of the United States of America, 94, 3817–3822.

Cheng, C.C.-H. (1998). Evolution of the diverse antifreeze proteins. Current Opinion in Genetics & Development, 8, 715–720.

Ciliberti, S., Martin, O.C., and Wagner, A. (2007a). Circuit topology and the evolution of robustness in complex regulatory gene networks. PLoS Computational Biology, 3(2), e15.

Ciliberti, S., Martin, O.C., and Wagner, A. (2007b). Innovation and robustness in complex regulatory gene networks. Proceedings of the National Academy of Sciences of the United States of America, 104, 13591–13596.

Cline, R.E., Hill, R.H., Phillips, D.L., and Needham, L.L. (1989). Pentachlorophenol measurements in body-fluids of people in log homes and workplaces. Archives of Environmental Contamination and Toxicology, 18, 475–481.

Conrad, M. (1990). The geometry of evolution. Biosystems, 24, 61–81.

Copley, S.D. (2000). Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. Trends in Biochemical Sciences, 25, 261–265.

Copley, R.R. and Bork, P. (2000). Homology among $(ba)_8$ barrels: Implications for the evolution of metabolic pathways. Journal of Molecular Biology, 303, 627–640.

Dantas, G., Sommer, M.O.A., Oluwasegun, R.D., and Church, G.M. (2008). Bacteria subsisting on antibiotics. Science, 320, 100–103.

Davidson, E.H. and Erwin, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. Science, 311, 796–800.

Eyre-Walker, A. and Keightley, P.D. (2007). The distribution of fitness effects of new mutations. Nature Reviews Genetics, 8, 610–618.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A. R., Karp, P. D., et al. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular Systems Biology, 3, 121.

Feist, A.M., Herrgard, M.J., Thiele, I., Reed, J.L., and Palsson, B.O. (2009). Reconstruction of biochemical networks in microorganisms. Nature Reviews Microbiology, 7, 129–143.

Feist, A.M. and Palsson, B.O. (2008). The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nature Biotechnology, 26, 659–667.

Ferrada, E. and Wagner, A. (2010). Evolutionary innovation and the organization of protein functions in sequence space. PLoS ONE, 5(11), e14172.

Forster, J., Famili, I., Fu, P., Palsson, B., and Nielsen, J. (2003). Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. Genome Research, 13, 244–253.

Futuyma, D.J. (1998). Evolutionary Biology. Sinauer, Sunderland, MA.

Gavrilets, S. (1997). Evolution and speciation on holey adaptive landscapes. Trends in Ecology & Evolution, 12, 307–312.

Gavrilets, S. (2005). Fitness landscapes and the origin of species. Princeton University Press, Princeton, NJ.

Gilbert, S.F. (1997). Developmental Biology. Sinauer, Sunderland, MA.

Giurumescu, C.A., Sternberg, P.W., and Asthagiri, A.R. (2009). Predicting phenotypic diversity and the underlying quantitative molecular transitions. PLoS Computational Biology, 5, e1000354.

Golding, G.B. and Dean, A.M. (1998). The structural basis of molecular adaptation. Molecular Biology and Evolution, 15, 355–369.

Goodman, M., Pedwaydon, J., Czelusniak, J., Suzuki, T., Gotoh, T., Moens, L., et al. (1988). An evolutionary tree for invertebrate globin sequences. Journal of Molecular Evolution, 27, 236–249.

Hafner, M., Koeppl, H., Hasler, M., and Wagner, A. (2009). "Glocal" robustness in model discrimination for circadian oscillators. PLoS Computational Biology, 5, e1000534.

Huang, W., Petrosino, J., Hirsch, M., Shenkin, P., and Palzkill, T. (1996). Amino acid sequence determinants of beta-lactamase structure and activity. Journal of Molecular Biology, 258, 688–703.

Hueber, S.D. and Lohmann, I. (2008). Shaping segments: Hox gene function in the genomic age. Bioessays, 30, 965–979.

Hughes, C.L. and Kaufman, T.C. (2002). Hox genes and the evolution of the arthropod body plan. Evolution & Development, 4, 459–499.

Huynen, M., Stadler, P., and Fontana, W. (1996). Smoothness within ruggedness: The role of neutrality in adaptation. Proceedings of the National Academy of Sciences of the United States of America, 93, 397–401.

Irish, V.F. (2003). The evolution of floral homeotic gene function. Bioessays, 25, 637–646.

Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K.N., et al. (2004). Dynamic control of positional information in the early Drosophila embryo. Nature, 430, 368–371.

Kauffman, S. and Levin, S. (1987). Towards a general-theory of adaptive walks on rugged landscapes. Journal of Theoretical Biology, 128, 11–45.

Kleina, L. and Miller, J. (1990). Genetic studies of the lac repressor. 13. Extensive amino-acid replacements generated by the use of natural and synthetic nonsense suppressors. Journal of Molecular Biology, 212, 295–318.

Lemons, D. and McGinnis, W. (2006). Genomic evolution of Hox gene clusters. Science, 313, 1918–1922.

Levitt, M. (2009). Nature of the protein universe. Proceedings of the National Academy of Sciences of the United States of America, 106, 11079–11084.

Liang, Y.H., Hua, Z.Q., Liang, X., Xu, Q., and Lu, G.Y. (2001). The crystal structure of bar-headed goose hemoglobin in deoxy form: The allosteric mechanism of a hemoglobin species with high oxygen affinity. Journal of Molecular Biology, 313, 123–137.

Lipman, D. and Wilbur, W. (1991). Modeling neutral and selective evolution of protein folding. Proceedings of the Royal Society of London Series B, 245, 7–11.

MacCarthy, T., Seymour, R., and Pomiankowski, A. (2003). The evolutionary potential of the Drosophila sex determination gene network. Journal of Theoretical Biology 225, 461–468.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 27, 29–34.

Provine, W.B. (1986). Sewall Wright and evolutionary biology. University of Chicago Press, Chicago, IL.

Raman, K. and Wagner, A. (2011). Evolvability and robustness in a complex signaling circuit Molecular Biosystems, 7, 1081–1092.

Rehmann, L. and Daugulis, A.J. (2008). Enhancement of PCB degradation by Burkholderia xenovorans LB400 in biphasic systems by manipulating culture conditions. Biotechnology and Bioengineering, 99, 521–528.

Reidys, C., Stadler, P., and Schuster, P. (1997). Generic properties of combinatory maps: Neutral networks of RNA secondary structures. Bulletin of Mathematical Biology, 59, 339–397.

Rennell, D., Bouvier, S., Hardy, L., and Poteete, A. (1991). Systematic mutation of bacteriophage T4 lysozyme. Journal of Molecular Biology, 222, 67–87.

Rodrigues, J.F. and Wagner, A. (2009). Evolutionary plasticity and innovations in complex metabolic reaction networks. PLoS Computational Biology, 5, e1000613.

Rodrigues, J.F. and Wagner, A. (2011). Genotype networks innovation, and robustness in sulfur metabolism. BMC Systems Biology, 5, 39.

Rost, B. (1997). Protein structures sustain evolutionary drift. Folding & Design, 2, S19–S24

Samal, A., Rodrigues, J.F.M., Jost, J., Martin, O.C., and Wagner, A. (2010). Genotype networks in metabolic reaction spaces. BMC Systems Biology, 4, 30.

Sanchez, L., Chaouiya, C., and Thieffry, D. (2008). Segmenting the fly embryo: logical analysis of the role of the Segment Polarity cross-regulatory module. International Journal of Developmental Biology, 52, 1059–1075.

Schilling, C.H., Edwards, J.S., and Palsson, B.O. (1999). Toward metabolic phenomics: Analysis of genomic data using flux balances. Biotechnology Progress, 15, 288–295.

Schuster, P. (2006). Prediction of RNA secondary structures: from theory to models and real molecules. Reports on Progress in Physics, 69, 1419–1477.

Schuster, P., Fontana, W., Stadler, P., and Hofacker, I. (1994). From sequences to shapes and back—a case-study in RNA secondary structures. Proceedings of the Royal Society of London Series B, 255, 279–284.

Searcoid, M.O. (2007). Metric spaces. Springer, London.

Sumedha, Martin, O.C., Wagner, A. (2007). New structural variation in evolutionary searches of RNA neutral networks. Biosystems, 90, 475–485.

Takiguchi, M., Matsubasa, T., Amaya, Y., and Mori, M. (1989). Evolutionary aspects of urea cycle enzyme genes. Bioessays, 10, 163–166.

Thornton, J., Orengo, C., Todd, A., and Pearl, F. (1999). Protein folds, functions and evolution. Journal of Molecular Biology, 293, 333–342.

Todd, A., Orengo, C., and Thornton, J. (1999). Evolution of protein function, from a structural perspective. Current Opinion in Chemical Biology, 3, 548–556.

Todd, A., Orengo, C., and Thornton, J. (2001). Evolution of function in protein superfamilies, from a structural perspective. Journal of Molecular Biology, 307, 1113–1143.

van der Meer, J.R. (1995). Evolution of novel metabolic pathways for the degradation of chloroaromatic compounds. Presented at Beijerinck Centennial Symposium on Microbial Physiology and Gene Regulation—Emerging Principles and Applications, The Hague, Netherlands, December.

van der Meer, J.R., Werlen, C., Nishino, S.F., and Spain, J.C. (1998). Evolution of a pathway for chlorobenzene metabolism leads to natural attenuation in contaminated groundwater. Applied and Environmental Microbiology, 64, 4185–4193.

von Dassow, G., Meir, E., Munro, E., and Odell, G. (2000). The segment polarity network is a robust development module. Nature, 406, 188–192.

Wagner, A. (2005a). Circuit topology and the evolution of robustness in two-gene circadian oscillators. Proceedings of the National Academy of Sciences of the United States of America, 102, 11775–11780.

Wagner, A. (2005b). Robustness and evolvability in living systems. Princeton University Press, Princeton, NJ.

Wagner, A. (2008). Neutralism and selectionism: A network-based reconciliation. Nature Reviews Genetics, 9, 965–974.

Wagner, A. (2011). The origins of evolutionary innovations. A theory of transformative change in living systems. Oxford University Press, Oxford.

Wagner, G.P., Amemiya, C., and Ruddle, F. (2003. Hox cluster duplications and the opportunity for evolutionary novelties. Proceedings of the National Academy of Sciences of the United States of America, 100, 14603–14606.

Wright, S. (1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proceedings of the Sixth International Congress on Genetics, 1, 355–366.