HANDBOOK OF SYSTEMS BIOLOGY

CONCEPTS AND INSIGHTS

Edited by Marian Walhout, Marc Vidal, & Job Dekker

Chapter 13

Genotype Networks and Evolutionary Innovations in Biological Systems

Andreas Wagner
University of Zurich, Institute of Evolutionary Biology and Environmental Studies, Y27-J-54, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Chapter Outline

INTRODUCTION

How new traits originate in life is a question that has occupied evolutionary biologists since Darwin's time. This holds especially for traits that are evolutionary innovations, i.e., qualitatively new features that benefit their carrier [1,2]. About this origin, the geneticist de Vries said in 1904 that Darwin's theory can explain the *survival* of the fittest but not its *arrival* [3]. Today, more than a century later, the biological literature contains many well-studied examples of innovations, fascinating case studies of natural history [1,4–10]. However, we still know little about any principles that might underlie the origin of innovations, other than the well-worn notion that a combination of mutation and natural selection may be necessary. We do not even know whether such principles exist. De Vries' statement makes clear that such principles would be principles of how biological systems bring forth novel and beneficial phenotypes. They would be principles of phenotypic variability.

To understand the origins of new phenotypes one needs to understand the relationship between genotype and phenotype. The genotype is the totality of an organism's genetic material. The phenotype is any other observable characteristic. It includes the morphology and behavior of complex organisms, the structure of cells, the expression pattern of genes and proteins, the biosynthetic abilities of an organism's metabolism, and the three-dimensional structure and function of its macromolecules, such as protein and RNA molecules.

New phenotypes often arise through mutations that alter an organism's genotype. Therefore, understanding phenotypic variability requires understanding how genotypic change translates into phenotypic change. Ideally, experimentation should provide this understanding [11,12]. However, a systematic understanding of the relationship between genotype and phenotype requires the analysis of thousands if not millions of different genotypes and their phenotypes. It is beyond reach of current experimental technologies for most systems. An alternative is to use existing comparative data about genotypes and their phenotypes. The necessary information is available only for a few kinds of system, for example proteins, where the structure and function of tens of thousands of proteins are available. In most other systems computational modeling of phenotypes will be essential for the foreseeable future. Fortunately, the tools of systems biology have allowed us to make great strides in such modeling. For example, within the last 15 years it has become possible to computationally predict the biosynthetic phenotypes of enormously complex metabolic networks comprising hundreds of enzymatic reactions [13,14]. The analyses reviewed in this chapter use such computational approaches, as well as comparative data and experimentation. Taken together,

Handbook of Systems Biology Concepts and Insights. http://dx.doi.org/10.1016/B978-0-12-385944-0.00013-7

these three lines of evidence point to a series of surprisingly simple principles behind life's ability to produce novel and beneficial phenotypes, i.e., its innovability.

Here, I will first devote three short sections to three central classes of systems and their phenotypes. Changes in these systems are the foundations of most, if not all evolutionary innovations. These system classes are metabolic networks, regulatory circuits, and molecules such as proteins and RNA. Subsequently, I will suggest how one can study phenotypic variability systematically in these system classes. The next section explains two fundamental concepts, that of a genotype space and the phenotypes therein, for these system classes. The two sections after that summarize recent evidence that these system classes share two organizational features of genotype space that facilitate phenotypic variability and evolutionary innovation. These are the existence of genotype networks (to be defined further below) and of a great phenotypic diversity in different neighborhoods of genotype space. The next section explains how these concepts can help explain the origins of evolutionary innovations. A final section suggests why system classes as different as these can share such similarities, and especially the existence of genotype networks. The reason is that systems in these classes typically operate in changing environments, which endows them with robustness to environmental change, but also to genetic change. The existence of genotype networks is a consequence of such robustness.

I emphasize that the principles I discuss here by no means negate the importance of other factors, such as environmental change, phenotypic plasticity, multifunctionality of biological systems, epigenetic change, gene duplication, and gradual evolution from simple to complex systems, for the ability to bring forth variable phenotypes [15–20]. The principles I discuss are complementary to other factors, and may even help clarify the role these factors play in phenotypic variability. They do not only apply to qualitatively new phenotypes, but also to beneficial quantitative changes in existing phenotypes — evolutionary adaptations, in the jargon of evolutionary biologists. A more comprehensive treatment can be found elsewhere [10].

METABOLIC NETWORKS AND THEIR INNOVATIONS

Large-scale metabolic networks are systems of hundreds to more than 1000 chemical reactions that are at work in every organism [21]. Their most fundamental task is to transform sources of chemical elements and energy into a chemical form that is useful to the organism. Evolutionary innovations in metabolism fall into multiple categories. An especially prominent category concerns traits that allow organisms to survive on new sources of food. Prokaryotes are the undisputed masters of such innovations. They are able to survive on sources of carbon and energy that are bizarre and toxic (to us), including methane, hydrogen gas, crude oil, antibiotics, and xenobiotic chemicals [22–24].

The likely reason why many metabolic innovations occur in prokaryotes is the ability of prokaryotes to exchange genes through a variety of mechanism [25,26]. Horizontal gene transfer can transform the genome of a prokaryote on short evolutionary timescales, such that even different strains of the same bacterial species may differ in hundreds of genes. Such horizontal gene transfer is the cause of many evolutionary innovations. A candidate example involves the prokaryote *Sphingomonas chlorophenolica*, which is able to metabolize the toxic xenobiotic compound pentachlorophenol. It does so through a sequence of four chemical reactions [27], none of which is new to *S. chlorophenolica*. Two of them are involved in degrading naturally occurring chlorinated compounds in other organisms. Two others are involved in the metabolism of the common amino acid tyrosine [27]. The innovation in *S. chlorophenolica*'s metabolism is the combination of these reactions. Such new combinations of reactions can be easily achieved through horizontal transfer of enzyme coding genes.

Prokaryotes may be the most prolific metabolic innovators, but metabolic innovations also occur in the evolution of higher, multicellular organisms. An example is the urea cycle, an innovation that occurred during the evolution of land-living animals. It allows animals to dispose of ammonia, a waste product of their metabolism that is toxic to cells, by converting it into urea that is excreted in urine. The urea cycle consists of five metabolic reactions, none of which are new to their carrier. Individually, they are widespread in many organisms. Four of these reactions are involved in the biosynthesis of arginine, and the fifth is involved in the degradation of arginine [28]. What is new is the combination of these five reactions into a metabolic cycle, a major innovation of biological waste management.

REGULATORY CIRCUITS AND THEIR INNOVATIONS

Regulation is a process that changes the activity of genes and their products. It can affect transcription, translation, post-translational modification, transport, as well as several other aspects of gene and protein function. Among all the known modes of regulation, transcriptional regulation is perhaps the most prominent [29–33]. The reason is that most modes of regulation ultimately affect the regulation of transcription. Transcriptional regulation is thus a backbone of regulatory processes inside an organism. Transcriptional regulation involves specialized proteins called transcription factors that bind regulatory DNA near a gene. The binding of one or more transcription factors to such regulatory DNA can activate or repress the transcription of a gene through interactions with the RNA polymerase that is responsible for transcribing the gene.

Regulation is often mediated by complex regulatory circuits. Such circuits consist of multiple molecules that mutually influence each other's activity. Transcriptional regulation is no exception. Transcription factors form regulatory circuits that can comprise dozens of proteins. These proteins regulate the transcription of the genes encoding them, and of many other genes downstream of the circuit genes [34–39]. In doing so, the circuit's proteins produce a gene expression pattern in which specific genes are activated or repressed, a state that can vary in space and time. A gene expression pattern is a transcriptional regulation circuit's phenotype. Such phenotypes play central roles in physiology and in embryonic development, the process that creates a viable adult organism from a fertilized egg [34,35,40].

Regulatory circuits in general, and transcriptional regulation circuits in particular, are involved in the evolution of many new traits. One example involves the evolution of eyespots on the back of butterfly wings [41–43]. These traits may help butterflies deter predators [41–43]. Eyespots start to form during development in regions that are called eyespot foci. These foci express the protein *Distal-less*, which is causally involved in eyespot formation. The number of eyespots that form on a wing corresponds to the number of regions that express *Distal-less* during early wing development. What is more, grafts of *Distal-less* expressing cells to developing wing tissue can be sufficient for eyespot formation in the graft's recipient [44]. *Distal-less* is a transcription factor, a member of a complex regulatory circuit with other functions in the development of wings and legs [41–43].

Another example involves the evolution of dissected leaves in plants [45, 46]. The ancestral leaves of flowering plants were most likely simple leaves, which have an undivided leaf blade [47]. Dissected leaves evolved from such simple leaves. In a dissected leaf, the leaf blade is subdivided into multiple smaller leaflets. Leaf dissection is a trait that may facilitate heat dissipation in hot terrestrial environments and help increase CO_2 uptake in water [45, 46]. Dissected leaves may have originated multiple times in the evolution of flowering plants [47]. During the development of dissected leaves, transcription factors of the KNOX (KNOTTED1-like homeobox) family play a crucial role. They are expressed in leaf primordials, which form close to the growing tip of a plant's shoot. Increasing the expression of KNOX genes during leaf formation can increase the number of leaflets that are forming; conversely, reducing their expression can reduce this number of leaflets [48]. KNOX genes are part of a regulatory circuit [48].

These are just two examples where regulatory proteins and the regulatory circuits they form are critically involved in the formation of an organism and its parts, as well as in the formation of a structure that was an innovation when it first became fully formed. Other prominent examples include the role of Hox genes in the formation of axial structures such as limbs in vertebrates, or the role of MADS box genes in the formation and diversification of flowers [7, 41–43].

MACROMOLECULES AND THEIR INNOVATIONS

Individual proteins and RNA macromolecules are not usually considered the subject of systems biology, but they should be. They are systems whose parts are amino acid or nucleotide monomers. These parts are strung together to form a whole macromolecule that folds intricately in three-dimensional space. Such macromolecules are responsible for all enzyme-catalyzed reactions that take place in a cell. They serve numerous other functions in addition, including transport, structural support, and communication, and they are behind numerous if not all new molecular functions that originated in life's history. Some of these functions involve very little change in a macromolecule's genotype. (This genotype is the DNA string that encodes the molecule, but for many purposes, a protein's amino acid sequence or an RNA molecule's nucleotide sequence can be viewed as the genotype.) An example of a new function requiring little change involves the enzyme l-ribulose-5-phosphate 4-epimerase from the bacterium *Escherichia coli*, which is necessary for *E. coli* to grow on the sugar arabinose as a carbon and energy source. A single amino acid change from histidine to asparagine at position 97 of this enzyme suffices to create a new enzymatic function, an aldolase that joins one molecule of dihydroxyacetone phosphate and one of glycoaldehyde phosphate [49]. New functions in other molecules require greater amounts of amino acid change. Take as an example antifreeze proteins. These proteins occur in numerous organisms that have to survive cold conditions, such as Arctic and Antarctic fish, as well as overwintering insects and plants. Antifreeze proteins lower the freezing point of an organism's body fluids. They originated multiple times independently, in different organisms, and sometimes rapidly, through multiple amino acid changes in various ancestor proteins [50–52]. These are just two examples of myriad evolutionary innovations that occurred in biological macromolecules.

All these three kinds of change – in metabolism, in regulatory circuits, and in macromolecules – may be involved in any one innovation, and in ways that are difficult to disentangle. It is nonetheless useful to study these kinds of change separately, in order to find out whether any commonalities exist among them. Such commonalities would point to more general principles of phenotypic variability.

TOWARDS A SYSTEMATIC UNDERSTANDING OF INNOVATION

By themselves, examples of innovations like those just discussed may not help us answer whether broader principles of innovation exist. To this end, it may be necessary to study innovation more systematically. One way to do that is to study a 'space' of possible innovation in each of the three system classes. This space is vast, too vast to understand exhaustively. But even by examining small samples of this space it is possible to learn about the structure of the entire space, and thus about principles of innovability. I will next discuss such a more systematic approach for each system class. Before that, however, I want to highlight three goals that a systematic understanding of innovation — an innovability theory — should achieve (others are highlighted elsewhere [10]).

The first goal reflects perhaps the most difficult problem that the origin of new beneficial trait poses to our understanding of evolution. Most mutations that affect an organism's genotype are deleterious, that is, their effects harm rather than benefit their carrier. Such mutations produce inferior phenotypes. Thus, to find new and *superior* phenotypes organisms may have to explore many mutant genotypes. At the same time, however, organisms need to preserve existing, well-adapted phenotypes. In other words, organisms have to be conservative and explore many new phenotypes at the same time. How both objectives can be achieved simultaneously is a question that a theory of innovation would have to answer.

The second goal relates to the observation that many innovations in the history of life have occurred more than once [6]. Dissected leaves may have evolved multiple times; so did antifreeze proteins; and so did many metabolic innovations. For example, life has solved the metabolic problem of incorporating atmospheric carbon (CO_2) into biomass at least three times in different ways, that is, through the Calvin–Benson cycle, through the reductive citric acid cycle, and through the hydroxypropionate cycle [53].

A third goal relates to the observation that some innovations seem to combine existing parts of a system to create a new function. I mentioned a metabolic pathway that can degrade pentachlorophenol, as well as the urea cycle, both of which involve new combinations of existing enzymatic reactions — parts of a metabolic system. Is this combinatorial nature of innovations a peculiarity of some innovations, or is it a more general phenomenon?

The framework I will discuss here suggests an answer to all three questions.

GENOTYPE SPACES AND THE PHENOTYPES THEREIN

This section discusses genotypes and phenotypes separately for each of the three systems classes mentioned above.

The metabolic genotype of an organism is the totality of its DNA that encodes metabolic enzymes. It is thus fundamentally a subset of its genome. While one can think of this genotype as a DNA string, is often more expedient to represent the genotype more compactly. Here is a compact representation that is well suited to study innovation systematically [54]. Consider the known universe of biochemical reactions, that is, chemical reactions catalyzed by an enzyme that are known to occur somewhere in some organism. This known universe of biochemical reactions currently comprises more than 5000 such reactions [55]. One can write these reactions as a list, as shown in Figure 13.1a, which represents each reaction by its stoichiometric equation. Any one organism, such as a human or the bacterium *E. coli*, will have enzymes that catalyze some of these reactions but not others. For reactions that are catalyzed in any one organism, write 1 next to the stoichiometric equation shown in Figure 13.1a. For every reaction that does not occur, write 0. The result of this procedure is a binary string that indicates which reactions do or do not take place in the metabolism of an organism. It is a compact description of a metabolic genotype, comprising all the enzymatic reactions that take place in a metabolic network. With this definition in mind, the notions of metabolic genotype and metabolic network are used here synonymously.

The totality of all possible metabolic genotypes — the set of all the binary strings defined above — constitutes a metabolic *genotype space*, a collection of possible metabolic genotypes. This space is vast, containing more than 2^{5000} possible genotypes, more metabolisms than could ever be realized on earth (and many of them surely useless to life as we know it). It is the space of all possible metabolisms that can be realized with a given set of biochemical reactions. To understand metabolism and metabolic innovation systematically is to understand the structure of this space, and the metabolic phenotypes that exist in it.

A few further concepts are useful in discussing metabolic genotype space. The first is that of a *neighbor*. Two metabolic networks are neighbors in genotype space if they differ in a single chemical reaction. A *neighborhood* of a metabolic network comprises all its 1-neighbors, all metabolic genotypes that differ from it in a single reaction. These concepts can be extended to k-neighbors, networks that differ from a given network in k chemical reactions. The distance of two metabolic networks indicates the fraction of reactions in which they differ. Two metabolic networks have a distance of $D = 0$ if they contain the same reactions; a distance of $D = 0.5$ if 50% of reactions that are catalyzed by one network are not catalyzed by the other

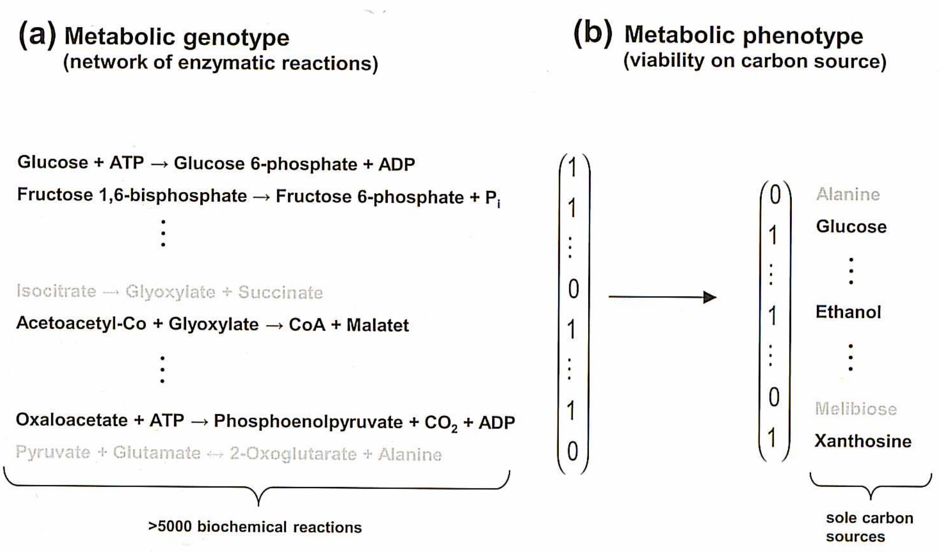
(a) Metabolic genotype
(network of enzymatic reactions)
Glucose + ATP → Glucose 6-phosphate + ADP
Fructose 1,6-bisphosphate → Fructose 6-phosphate + P$_i$
Isocitrate → Glyoxylate + Succinate
Acetoacetyl-Co + Glyoxylate → CoA + Malatet
Oxaloacetate + ATP → Phosphoenolpyruvate + CO$_2$ + ADP
Pyruvate + Glutamate ↔ 2-Oxoglutarate + Alanine
>5000 biochemical reactions
1
1
0
1
1
0
(b) Metabolic phenotype
(viability on carbon source)
0 Alanine
1 Glucose
1 Ethanol
0 Melibiose
1 Xanthosine
sole carbon sources

FIGURE 13.1 Metabolic genotypes and phenotypes. Panel a) shows the metabolic genotype of a genome-scale metabolic network. It can be represented in a simplified form as a binary string. The entries of this string correspond to one biochemical reaction in a 'universe' of known reactions. Panel b) shows one of many possible representations of a metabolic phenotype, a binary string representation whose entries correspond to individual carbon sources. This representation contains a one for every carbon source from which a metabolic network can synthesize all biomass precursors. *(Figure and caption adapted from [10]. Used with permission from Oxford University Press.)*

network or vice versa; and a distance of $D = 1$ if they differ in every single reaction [54,56,57].

There are as many ways to define a metabolic *phenotype* as there are tasks of metabolism. Metabolism detoxifies waste, synthesizes molecules for defense and communication, and manufactures all small precursor molecules for biomass synthesis. The latter task is the most fundamental, and I will therefore focus on metabolic phenotypes related to this task. For free-living organisms such as *E. coli* there are of the order of 60 small biomass precursors [58]. These include all proteinaceous amino acids, DNA nucleotide precursors, RNA nucleotide precursors, as well as multiple lipids and enzyme cofactors. A network's ability to synthesize all these molecules will depend on the nutrients that are available in an environment. Some organisms, such as *E. coli*, can survive in very simple, minimal chemical environments. These environments contain only one kind of molecule that provides each chemical element; at least one of these molecules also provides energy. Most free-living organisms can use multiple different sources of chemical elements and of energy.

These observations give rise to the following definition of a metabolic phenotype, which is focused on sources of carbon and energy but can be easily extended to sources of other chemical elements [54,57]. Consider a given number of molecules that could serve as sources of carbon energy to some organism. Write these molecules as a list, as shown in Figure 13.1b. If the metabolism of a given organism can synthesize biomass — that is, if it can sustain life on any one of these carbon sources (that is, the organism needs to be able to use this carbon source as its *only* carbon source) — write a 1 next to the carbon source. Otherwise, write a 0. In this way one can define a metabolic phenotype as a binary string that reflects an organism's viability on different sources of carbon or other elements. Note that even for a modest number of 100 different potential carbon sources, the number of possible metabolic phenotypes is already 2^{100} or $>10^{30}$.

This representation of metabolic phenotypes lends itself to the systematic study of new metabolic phenotypes. Consider a genotypic change that causes the addition of chemical reactions to a metabolic network by horizontal gene transfer. If these new reactions allow an organism to survive on a carbon source that it had not been previously able to utilize, a metabolic innovation has arisen. In an environment where other carbon sources limit growth, or where they are absent, this ability can make a life-changing qualitative difference to its carrier.

I will now discuss analogous definitions of genotypes and phenotypes for regulatory circuits. The evolution of regulatory circuits, and especially of transcriptional regulation circuits, is difficult to study experimentally. Part of the reason is that regulatory DNA can occur far away from the genes it regulates; also, such DNA can change very rapidly on evolutionary timescales [36,59–66]. In addition, to understand the relationship between genotype and phenotype requires an analysis of many circuit genotypes and their phenotypes. For these reasons, computational models of regulatory circuits are still indispensable to understand genotype–phenotype relationships in such circuits. The evidence discussed below stems from well-studied models of transcriptional regulation circuits [67–71]. Variants of these models have been used successfully to understand the development of specific organisms, such as the early fruit fly embryo, and to predict the developmental changes in mutant embryos [67, 72–75]. In addition, they have helped us understand a variety of evolutionary phenomena, such as how regulatory circuits can evolve increased robustness to perturbations, and that cryptic variation — genotypic variation without phenotypic effects in a given environment — might facilitate evolutionary adaptation [76–80]. Note that circuits different from those discussed here, such as signaling circuits, can show properties similar to those highlighted below, which suggests that these properties may be generic features of regulatory circuits [71,81,82].

The genotype of a regulatory circuit comprises the genomic DNA that encodes all regulatory molecules, as well as the non-coding DNA that may help determine the interactions between them. For a transcriptional regulation circuit, this genotype typically includes the genes that encode the circuit's transcriptional regulators, as well as the regulatory DNA sequences that determine where a regulator binds, and which therefore determine who regulates who in the circuit. As in metabolism, there are more compact representations of a circuit's regulatory genotype than its DNA sequence. For example, one can represent the regulatory genotype simply through a square matrix $w=(w_{ij})$, whose entries w_{ij} reflect whether transcription factor j regulates the expression of transcription factor i in the circuit. In the simplest possible representation, this matrix contains only information about whether this interaction is activating ($w_{ij}=+1$), repressing ($w_{ij}=-1$), or absent. Even the simplest representation shows that the number of circuit genotypes will be very large, even for circuits with a modest number N of genes. That is, there are 3^{N^2} possible circuits. In more complicated representations of transcriptional regulation circuits, these interactions could assume a larger or a continuous range of values. Mutations in DNA that affect the regulatory interactions of circuit genes can change this circuit genotype. For example, a mutation in regulatory DNA that abolishes binding of a transcription factor to this DNA may also abolish regulation of a nearby gene by this transcription factor, and thus eliminate one of the regulatory interactions w_{ij} of this circuit ($w_{ij}\rightarrow 0$).

The mutual regulatory interactions of molecules in such a circuit will create a gene expression pattern. This expression pattern is a circuit's phenotype. It typically influences the expression of many genes downstream of the circuit, genes that influence physiological or developmental processes. Changes in such phenotypes caused by mutations of the circuit's regulatory genotype can help create new traits, some of which may become evolutionary innovations.

To study the origin of new gene expression phenotypes in such circuits systematically, one needs to think of any one circuit as being part of a much larger genotype space of circuits. This space contains all possible circuits of a given number N of genes. In this genotype space, two circuits are *k-neighbors* if they differ in k regulatory interaction. The *k-neighborhood* of a circuit comprises all circuits that differ from it in no more than k regulatory interaction. The distance D of two circuits can be defined as the fraction of regulatory interactions in which they differ. For example, two circuits would have a distance of $D=0.2$, if they differed in 20% of their interactions. They would have a distance of $D=1$ if they differed in every single interaction, that is, if no interaction that occurs in the first circuit also occurs in the second circuit [83,84].

The final class of systems to be discussed here are protein and RNA macromolecules. Their genotype spaces, also known as sequence spaces, have been studied for many years [85–87]. For protein strings of a given length N of amino acids, genotype space comprises all amino acids strings of length N, and thus a totality of 20^N such strings, because 20 different amino acids occur in most proteins. For RNA molecules of N nucleotides, it comprises 4^N possible RNA strings. As for metabolism and for regulatory circuits, the sizes of these genotype spaces can be astronomically large. Two protein and RNA molecules are k-neighbors in genotype space if they differ in k nucleotides or amino acids. The k-neighborhood of a molecule comprises all of its neighbors. The distance of two protein or RNA molecules can be defined in a variety of ways, one of them being the fraction of monomers in which they differ.

The phenotype of a protein or RNA molecule comprises its secondary structure, its tertiary structure — that is, its three-dimensional fold in space — as well as its biochemical function, be it catalytic, structural, or something else. Over the last 40 years the genotypes and phenotypes of tens of thousands of proteins have been characterized biochemically. They provide a rich source of information to study the relationship between genotype and phenotype [88]. Fewer RNA phenotypes are known, but for RNA secondary structures algorithms exist that can predict RNA phenotypes from genotypes [89,90]. Albeit not perfectly accurate for any one sequence, the relevant algorithms are sufficiently accurate (and also sufficiently fast) to characterize thousands to millions of different RNA genotypes and their phenotypes [87,91–93]. Because RNA secondary structure phenotypes are necessary for the functioning of many molecules, they are interesting study objects in their own right [94–96].

Genotype Networks

The genotype spaces of metabolic networks, regulatory circuits, and macromolecules are much too large to be characterized exhaustively. However, they can be characterized through (unbiased) sampling of genotypes or phenotypes, or through exhaustive enumeration of genotypes and phenotypes for small systems. These approaches can identify generic properties of such spaces, that is properties that hold for typical genotypes and phenotypes

Two such properties are discussed in this and the following section.

The first is that a given phenotype is typically not just formed by one or few genotypes, but by astronomically many genotypes [54,56,84,86,87]. In other words, vast sets of genotypes share the same phenotype. In some systems, such as metabolism, it is possible to characterize such sets of genotypes through Markov chain Monte Carlo sampling of genotype space [54,56]. This involves carefully designed random walks through genotype space. Briefly, one starts from a specific metabolic network (metabolic genotype) with a given number of reactions and a given phenotype. (This starting genotype can be viewed as a single point in metabolic genotype space.) Techniques such as flux-balance analysis allow one to compute this metabolic phenotype from the network's metabolic genotype [14,97]. One then either eliminates a specific, randomly chosen reaction from this starting genotype, or one adds a reaction chosen at random from the known universe of biochemical reactions. After this change to the network, one computes the phenotype of the changed network. If this phenotype is the same as before the change — that is, if this change has not altered the ability of the network to sustain life on a given spectrum of carbon sources — then the altered network is kept. Otherwise, the genotypic change is discarded and one reverts to the initial metabolic network. One then applies a second change (reaction deletion or addition), evaluates the phenotype, and keeps the altered network if it is unchanged, and so on, in a long sequence of $>10^5$ reaction changes, each of which has to keep the network's phenotype unchanged. This approach can not only sample sets of genotypes with a given phenotype uniformly, that is, in an unbiased manner, it also resembles the process by which metabolic networks evolve through the deletion and the addition of reactions to a network, for example through horizontal gene transfer.

Using this approach one can ask how different two metabolic genotypes that have the same phenotype can become? The answer is that they can become very different. For example, metabolic networks that have the same number of reactions as the *E. coli* network, and that can synthesize all *E. coli* biomass precursors in a minimal environment that contains glucose as the *sole* carbon source, can differ in more than 75% of their reactions. Moreover, any two such metabolic networks can typically be connected to one another. This means that sequences of reaction changes exist that can convert one metabolic network into the other, such that no individual change alters the phenotype [54,56]. In other words, metabolic genotypes with the same phenotype form extended networks — genotype networks — in metabolic genotype space. Note the distinction between a metabolic network and a genotype network: a metabolic network corresponds to a single point, a single genotype in genotype space; a genotype network is a network of such genotypes, and thus a network of metabolic networks. I will keep these two meanings of a network distinct.

Genotype networks exist for metabolic phenotypes that can sustain life on many different sole carbon sources, as well as on multiple carbon sources (when each source is provided as the sole carbon source). Even metabolic networks that can sustain life on up to 60 carbon sources form genotype networks that can still differ in 75% of their reactions. Extended genotype networks also exist for metabolic networks that contain different numbers of reactions, and for phenotypes that involve sources of chemical elements other than carbon [57]. Thus, the most basic properties of genotype networks are not highly sensitive to the phenotype one considers. Based on what we know, they appear to be generic features of metabolic genotype space.

To explore the genotype space of regulatory circuits one can use similar sampling approaches [83,84], and one finds a similar organization of this space. Two circuits with the same gene expression phenotype can have a genotype distance between $D=0.75$ and $D=1$, that is, they may differ in 75—100% of their regulatory interactions. What is more, circuits with very different genotypes can typically be connected through a sequence of steps, each of which changes a single regulatory interaction, but none of which alters the circuit's gene expression phenotype. These observations hold for broadly different gene expression phenotypes, and regardless of the number of genes or regulatory interactions in a circuit, except possibly for the smallest circuits [84,98].

Comparative data on the tertiary structure phenotypes of proteins demonstrates the existence of genotype networks here as well. Although exceptions exist [99], proteins with the same structure and/or function can differ in most of their amino acids [100—103]. Examples include oxygen-binding globins. These proteins occur both in animals and plants, probably share a common ancestor, but are extremely diverse in their genotypes. For example, no more than four of their more than 90 amino acids are absolutely conserved. Despite this genotypic divergence, globins have largely preserved their tertiary structure and their oxygen-binding ability [104—106].

Globins are not unusual in this regard. Other proteins with preserved phenotype are even more diverse. Take triose phosphate isomerase (TIM) barrel proteins. These proteins have preserved their tertiary structure but can differ in every single amino acid [107,108]. More generally, proteins with highly diverged genotype yet highly conserved phenotype are the rule rather than the exception [100—103]. Such proteins form vast genotype networks that extend far into genotype space. Phylogenetic analyses of related proteins from different organisms reveal a reflection of these networks in the tree of life [106].

Fewer RNA than protein phenotypes have been characterized experimentally, but computational analyses of the relationship between RNA sequence and secondary structure point to much the same phenomenon. RNA genotypes with the same secondary structure phenotype typically form large genotype networks that extend far into, and often all the way through, RNA genotype space [87,109].

In sum, metabolic networks, regulatory circuits, and macromolecules — very different kinds of systems — show a remarkable common property. Genotypes that have the same phenotype are typically organized in large genotype networks that reach far through genotype space. I will return to genotype networks later, when I discuss their significance for phenotypic variability.

The Diversity of Neighborhoods in Genotype Space

A second common property of the three system classes emerges from the analysis of genotypic neighborhoods. The neighborhood of a genotype is relevant for phenotypic variability, because it contains genotypes that can be easily reached from this genotype, that is, through one or few small genotypic changes. For an analysis of phenotypic variability, it is therefore useful to examine the spectrum of phenotypes ***P1*** that occur in a given neighborhood of a genotype *G1* that has some phenotype *P*. A simple question is whether the spectrum of phenotypes in this neighborhood depends on the genotype *G1*. More precisely, consider two genotypes *G1* and *G2* with the same phenotype *P* and a given distance *D*. Denote as ***P1*** and ***P2*** the sets of phenotypes (different from *P*) in their respective neighborhoods. How different is the set ***P1*** from the set ***P2***? That is, are most phenotypes in ***P1*** also contained in ***P2***? Or are most of these phenotypes unique to the neighborhood of *G1*, in the sense that they do not also occur in the neighborhood of *G2*?

In metabolism, one finds that the neighborhoods of two metabolic genotypes *G1* and *G2* sampled at random from the same genotype network contain mostly different novel phenotypes. In other words, the set ***P1*** of new phenotypes in the neighborhood of *G1* is very different from the set ***P2*** of new phenotypes in the neighborhood of *G2*. This holds regardless of the specific genotypes *G1* and *G2*, as well as regardless of the specific phenotype *P* that they have [54,57]. The situation in regulatory circuits is not much different. There, small neighborhoods around two circuits *G1* and G2 may contain sets of phenotypes ***P1*** and ***P2*** that differ in the majority of their phenotypes, even for circuits whose genotypes differ little, that is, in no more than 20% of their regulatory interactions [83]. Much the same holds for protein and RNA molecules [87,92,110,111]. For example, a recent analysis studied more than 16 000 enzymes with known sequence, tertiary structure, a enzymatic function. It showed that small neighborho around two proteins *G1* and *G2* that differ at fewer th 25% of their amino acids can contain sets of ***P1*** and ***P2*** new enzyme function phenotypes, such that the majority enzymatic functions found in ***P1*** are not contained in ***P***

In sum, metabolic networks, regulatory circuits, a macromolecules show two common qualitative propert in the organization of their genotype space. Property 1 the existence of genotype networks that reach far throu genotype space. Property 2 is that small neighborho around different genotypes typically contain differ phenotypes, even if the genotypes do not differ grea Figure 13.2 shows a schematic sketch of these properti The large rectangle in the figure stands for a hypotheti genotype space. Each of the small circles stands for a sin genotype. The open circles correspond to genotypes t share some hypothetical phenotype *P* (not show Two genotypes are connected by a straight line if they

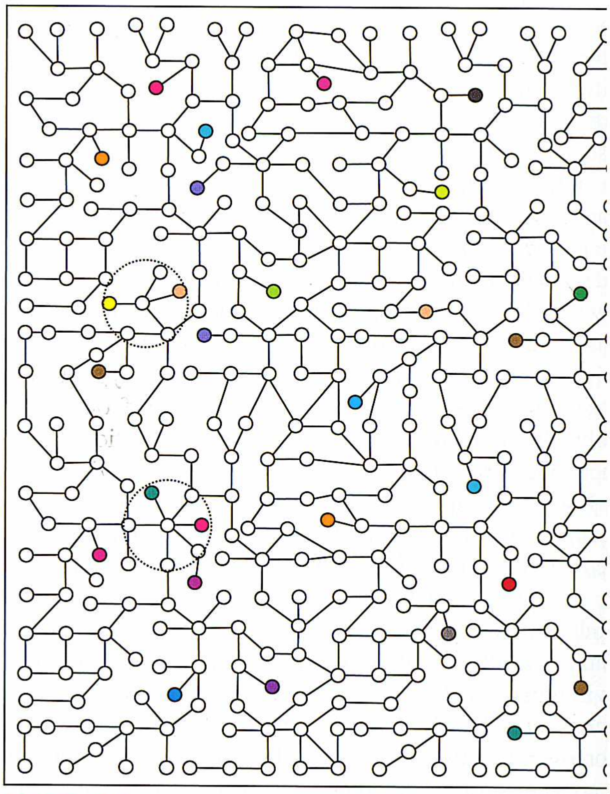

FIGURE 13.2 **A highly simplified schematic of the structure a genotype network and the new phenotypes near it. See text details.** Note that genotype networks are objects in a high-dimensio genotype space with counterintuitive geometric properties. Also, act genotype networks contain an astronomical number of members. In vidual genotypes may have hundreds to thousands of neighbors, only f of which can be shown. In addition, each of the genotypes shown different colors is also part of a vast genotype network that is not shown figure like this can thus merely provide a modicum of intuition about organization of genotype space. (*Adapted from [10]. Used with permissi from Oxford University Press.*)

neighbors. The network of open circles stands for a large connected genotype network that traverses genotype space. The colored circles represent genotypes whose phenotype is different from P (one color per phenotype), and that are neighbors of the genotypes on the genotype network. Note that different regions of this hypothetical genotype space contain different colors. The two large dashed circles represent the neighborhoods of two genotypes on the genotype network. Note that the phenotypes (colors) in these two neighborhoods are different, a reflection of property 2. Note that Figure 13.2 represents a complex, vast, and high dimensional genotype space in a highly simplified, two-dimensional way. For example, actual genotype networks contain an astronomical number of members. Individual genotypes may have hundreds to thousands of neighbors, only few of which can be shown. In addition, each of the colored genotypes is also part of a vast genotype network that is not shown.

Genotype Networks and Their Diverse Neighborhoods Can Help Explain the Origin of New Phenotypes

I will now return to the three questions about the origin of new phenotypes posed earlier, and which a systematic understanding of phenotypic variability needs to address.

The first is that organisms need to preserve old, well-adapted phenotypes while exploring many new phenotypes. Properties 1 and 2 can jointly help answer this question. To see this, consider that all evolution takes place in populations of organisms, each with its own genotype. Envision a population of genotypes in any one of our three system classes. Individuals in this population have a phenotype that may be necessary for their survival, but somewhere in genotype space a superior phenotype may exist. The genotypes of individuals in this population suffer mutations that affect their genotype. Natural selection eliminates any mutants that have not preserved the old phenotype or replaced it with a superior phenotype. One can view such a population as a cloud of points [112] that diffuses on a genotype network through genotype space.

Genotype networks (property 1) allow the genotypes of individuals in such a population to change without affecting their phenotype. They allow the preservation of old phenotypes despite genotypic change. Over time, genotypes may change dramatically while preserving their phenotype. During this process, the population explores different regions of genotype space. Because of property 2, the diversity of genotypic neighborhoods, the neighborhoods of the population's genotypes will contain ever-changing sets of new phenotypes. This means that the population can explore different novel phenotypes in its neighborhood as its genotypes change. In sum, genotype networks and their neighborhoods allow populations to preserve old phenotypes while exploring many new phenotypes.

Neither property 1 nor property 2 alone would be sufficient for such exploration [10]. Without property 1 (no genotype networks), a population would have low genotypic diversity and could therefore not explore different neighborhoods in this space. The total number of phenotypes is much greater than the number of phenotypes in a neighborhood for any one of the three system classes [10]. Thus, the absence of genotype networks would mean that most novel phenotypes are off-limits to an evolving population. Conversely, in the absence of property 2, that is, if the neighborhoods of different genotypes contained mostly identical new phenotypes, the existence of genotype networks would be irrelevant to the exploration of novel phenotypes. The reason is that even though a population's genotypes could change during evolutionary exploration of a genotype network, the changing genotypes would have access to the same unchanging spectrum of novel phenotypes.

The second question posed earlier regards the multiple evolutionary origins of many evolutionary innovations [6, 53]. Such multiple origins may be difficult to understand, if one assumes that innovations are unique solutions to particular problems that life faces, and that they are unique because the underlying problems are difficult to solve. Viewing such solutions from the vantage point of a genotype space leads to a completely different perspective. There, a genotype with a specific phenotype can be viewed as a solution to a particular problem. The existence of vast genotype networks for typical phenotypes means that typical problems have not just one, but astronomically many solutions. Different genotypes on the same genotype network can be viewed as different solutions to the same problem. Populations of organisms that explore genotype space from different starting points may encounter different solutions. To be sure, most innovations may involve multiple changes in all three major system classes, but because genotype networks are ubiquitous in all three classes, so are multiple solutions to most problems. From this perspective, the multiple origins of many evolutionary innovations are not surprising but rather to be expected.

The third question is whether innovation is usually combinatorial, involving old parts that are combined to new purposes. Here again, the vantage point of a genotype space, which contains all possible innovations, suggests a very straightforward answer: all innovation is combinatorial. New functions of proteins emerge through new combinations of amino acids. New metabolic phenotypes emerge through new combinations of already existing biochemical reactions. And new gene expression patterns of regulatory circuits arise through new combinations of regulators and their interactions.

Innovations that involve new combinations of existing system parts have many guises. Students of embryonic development, for example, have coined the term co-option, the use of an existing regulator or an existing regulatory interaction for new purposes [113]. Examples include the regulator *Distal-less* mentioned earlier. It is involved in the development of insect legs and wings, but it has been co-opted to form eyespots [44,114]. *Distal-less* does not act alone in these processes. It is part of regulatory circuit that involves other molecules, some of which may also have changed their interactions and expression in helping form a new body structure. One can view the co-option of *Distal-less* as a special case of a more general principle, in which new combinations of regulators and their interactions specify new body parts.

In sum, evidence from three very different kinds of systems can help answer several related questions about the origins of new phenotypes. It can help us grasp how life can preserve old phenotypes while exploring many new phenotypes. It can help us understand how many evolutionary innovations have originated multiple times in the history of life. And it can help us appreciate that innovation will generally involve combinations of old parts to achieve new purposes. The fact that the properties described exist in very different kinds of system suggests that they apply to multiple different kinds of innovation. They are suitable to form the basis of a general innovability theory.

Robustness, Genotype Networks and Environmental Change

A question so far left open is why genotype networks exist in metabolism, regulatory circuits, and macromolecules. At first sight this may seem difficult to answer, because these system classes are so different. However, it can be shown that this commonality emerges from a very simple property that they share: the robustness of their phenotypes to mutational changes in individual system parts.

In the genotype space framework such robustness can be thought of as a property of individual genotypes. Mutations often change any one genotype into one of its neighbors. A loss-of-function mutation in an enzyme-coding gene may eliminate one reaction from a metabolic network and transform the network into one of its neighbors; a mutation-changing regulatory DNA may eliminate a transcription factor's binding to this DNA, and hence its regulatory interaction with a target gene, transforming the circuit into one of its neighbors; a nucleotide change in a protein-coding gene often transforms the protein into one of its neighbors. One way to quantify the robustness of a genotype is through the proportion of its neighbors that have the same phenotype as itself. Metabolic networks, regulatory circuits and macromolecules are all to some extent robust in this sense [39,115—123]. This robustness has been estimated experimentally in systems such as proteins through random mutagenesis experiments [115—118], in metabolic networks through knockout mutations of enzyme-coding genes, and in regulatory circuits through circuit rewiring [39,115—123]. Computer modeling confirms that such robustness is a generic feature of these three system classes [54,84,124]. Typically, between 10% and more than 50% of a genotype's neighbors have the same phenotype as itself, depending on the system and the individual genotype [10]. It can be shown mathematically that this property is both necessary and sufficient to bring forth genotype networks that are astronomically large, and that extend far through genotype space [10]. From this vantage point, one could argue that genotype networks are a consequence of robustness. (Their diverse phenotypic neighborhoods emerge from the fact that many more phenotypes exist than the neighborhood of any one genotype can contain [10].)

These observations raise a further question. What is the ultimate cause of this robustness? Although multiple answers have been proposed, the current best candidate emerges from the observation that living systems need to operate in different environments [119,125—129]. The notion of an environment should be broadly defined in this context, and include the biotic, chemical, and physical environment outside an organism, as well as inside its cells. For example, it includes the changing chemical environments that provide nutrients to a metabolic network, the different regions of a developing embryo in which a regulatory circuit is exposed to different chemical signals, and the intracellular chemical environment that macromolecules need to operate in.

The role of changing environments for robustness has been most thoroughly studied in the context of metabolic networks [56, 119,129—133]. A free-living organism such as *E. coli*, which encounters multiple different environment containing different nutrients, can sustain life on dozens of different nutrients. It also has a large metabolic network that comprises more than 900 reactions. In any one such environment, it is also robust to the removal of individual reactions [134]. For example, more than 70% of its reactions are dispensable in a minimal environment with glucose as the sole carbon source. Such robustness is not a peculiarity of the *E. coli* metabolic network: it is a general property of metabolic networks that can sustain life in multiple environments [57,119]. (Note that any reaction that is dispensable in one environment may be essential in a different environment [119].) If *E. coli* lived for many generations in an environment that did not vary in its nutrient composition, its robustness would slowly disappear. This is what happened in endosymbiotic organisms such as *Buchnera aphidicola*, a relative of *E. coli* that has lived for millions of years inside its host organism, an aphid

[135–137]. *Buchnera* has a much simpler metabolic network comprising only 263 reactions. During its long association with its host and the constant environment this host provides, *Buchnera* has lost the ability to survive on a broad spectrum of nutrients. What is more, it has also lost almost all robustness to the removal of chemical reactions from its metabolism [135].

A metabolism that is to sustain life in multiple different chemical environments needs specific enzymatic reactions to metabolize all the nutrients that these environments may contain. It therefore needs to be more complex than a metabolism highly specialized to one specific environment. This increased complexity endows metabolism with robustness to the removal of reactions in any one environment [10,57]. Although the relationship between environmental change, increased complexity, and robustness is not as well explored for proteins and regulatory circuits, similar arguments can be made for them [10].

In sum, the ability to cope with changing environments can require increased complexity of a biological system, which can cause robustness to genetic change in any one environment. Such robustness is responsible for the existence of genotype networks, which can facilitate the origin of new phenotypes. Minimally complex systems may not display some of the core properties discussed here [98], and may thus not be capable of exploring a broad spectrum of new phenotypes.

Conclusions and Future Challenges

Genotypic changes in metabolism, in regulatory circuits, as well as in protein and RNA molecules are involved in many if not all evolutionary innovations. These system classes are therefore important study objects to understand the phenotypic variability that brings forth evolutionary adaptations and innovations. I have discussed evidence that all three system classes have two common properties. The first is the existence of genotype networks, vast connected sets of genotypes with the same phenotype that reach far through genotype space. The second is the fact that the neighborhoods of different genotypes with the same phenotype typically contain very different new phenotypes. Together, these properties can help explain how living systems can preserve old phenotypes while exploring many new phenotypes, why many evolutionary innovations in life's history have occurred multiple times, and that evolutionary innovation has a fundamentally combinatorial nature. Robustness of a system to genetic change is both a necessary and sufficient criterion for the existence of genotype networks. A likely cause of such robustness is the fact that many biological systems need to operate in multiple environments.

The observations made here regard qualitative commonalities in the organization of genotype spaces. These spaces may harbor further, unrecognized similarities, but also differences among different kinds of systems. A combination of high-throughput genotyping with emerging technologies for high-throughput phenotyping [138], and sophisticated computer models of genotype–phenotype relationships may reveal many more such principles in the years to come. Given how vast genotype spaces are, myriad principles of phenotypic variability may still await discovery.

REFERENCES

[1] Muller GB, Wagner GP. Novelty in evolution – restructuring the concept. Annu Rev Ecol Evol Syst 1991;22:229–56.

[2] Pigliucci M. What, if anything, is an evolutionary novelty? In 20th Biennial meeting of the philosophy of science association. CANADA: Vancouver; 2006. p. 887–98.

[3] de Vries H. Species and Varieties, Their Origin by Mutation. Chicago, IL: The open court publishing company; 1905.

[4] Shubin N, Tabin C, Carroll S. Deep homology and the origins of evolutionary novelty. Nature 2009;457:818–23.

[5] Moczek AP. On the origins of novelty in development and evolution. Bioessays 2008;30:432–47.

[6] Vermeij GJ. Historical contingency and the purported uniqueness of evolutionary innovations. Proc Natl Acad Sci USA 2006;103:1804–9.

[7] Irish VF. The evolution of floral homeotic gene function. Bioessays 2003;25:637–46.

[8] Shimeld SM, Holland PWH. Vertebrate innovations. Proc Natl Acad Sci USA 2000;97:4449–52.

[9] Gerhart J, Kirschner M. Cells, Embryos, and Evolution. Boston: Blackwell; 1998.

[10] Wagner A. The Origins of Evolutionary Innovations. A Theory of Transformative Change in Living Systems. Oxford, UK: Oxford University Press; 2011.

[11] Schultes E, Bartel D. One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science 2000;289:448–52.

[12] Hayden E, Ferrada E, Wagner A. Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. Nature 2011;474:92–5.

[13] Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO. Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 2009;7:129–43.

[14] Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protocol 2007;2:727–38.

[15] West-Eberhard M. Developmental Plasticity and Evolution. New York, NY: Oxford University Press; 2003.

[16] Jeffery C. Moonlighting proteins. Trends Biochem Sci 1999;24:8–11.

[17] True HL, Berlin I, Lindquist SL. Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. Nature 2004;431:184–7.

[18] Masel J, Bergman A. The evolution of the evolvability properties of the yeast prion [PSI+]. Evolution 2003;57:1498–512.

[19] Jensen RA. Enzyme recruitment in evolution of new function. Annu Rev Microbiol 1976;30:409–25.

[20] Horowitz NH. On the evolution of biochemical syntheses. Proc Natl Acad Sci USA 1945;31:153–7.

[21] Palsson B. Metabolic systems biology. FEBS Lett 2009;583: 3900–4.

[22] Postgate JR. The Outer Reaches of Life. Cambridge, UK: Cambridge University Press; 1994.

[23] Dantas G, Sommer MOA, Oluwasegun RD, Church GM. Bacteria subsisting on antibiotics. Science 2008;320:100–3.

[24] Nohynek LJ, Suhonen EL, NurmiahoLassila EL, Hantula J, SalkinojaSalonen M. Description of four pentachlorophenol-degrading bacterial strains as *Sphingomonas chlorophenolica* sp nov. Syst Appl Microbiol 1996;18:527–38.

[25] Ochman H, Lawrence J, Groisman E. Lateral gene transfer and the nature of bacterial innovation. Nature 2000;405:299–304.

[26] Pal C, Papp B, Lercher MJ. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat Genet 2005;37:1372–5.

[27] Copley SD. Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. Trends Biochem Sci 2000;25:261–5.

[28] Takiguchi M, Matsubasa T, Amaya Y, Mori M. Evolutionary aspects of urea cycle enzyme genes. Bioessays 1989;10:163–6.

[29] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. New York, NY: Garland Science; 2008.

[30] Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, et al. Diversity and Complexity in DNA Recognition by Transcription Factors. Science 2009;324:1720–3.

[31] Walhout AJM. Unraveling transcription regulatory networks by protein-DNA and protein–protein interaction mapping. Genome Res 2006;16:1445–54.

[32] Davidson EH, Erwin DH. Gene regulatory networks and the evolution of animal body plans. Science 2006;311:796–800.

[33] Arnone MI, Davidson EH. The hardwiring of development: organization and function of genomic regulatory systems. Development 1997;124:1851–64.

[34] Hueber SD, Lohmann I. Shaping segments: Hox gene function in the genomic age. Bioessays 2008;30:965–79.

[35] Hughes CL, Kaufman TC. Hox genes and the evolution of the arthropod body plan. Evol Dev 2002;4:459–99.

[36] Tuch BB, Li H, Johnson AD. Evolution of eukaryotic transcription circuits. Science 2008;319:1797–9.

[37] Lee T, Rinaldi N, Robert F, Odom D, Bar-Joseph Z, Gerber G, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 2002;298:799–804.

[38] Shen-Orr S, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat Genet 2002;31:64–8.

[39] Isalan M, Lemerle C, Michalodimitrakis K, Beltrao P, Horn C, Garriga-Canut M, et al. Evolvability and hierarchy in rewired bacterial gene networks. Nature 2008;452:840–5.

[40] Carroll SB, Grenier JK, Weatherbee SD. From DNA to Diversity. Molecular Genetics and the Evolution of Animal Design. Malden, MA: Blackwell; 2001.

[41] Stevens M, Stubbins CL, Hardman CJ. The anti-predator function of 'eyespots' on camouflaged and conspicuous prey. Behav Ecol Sociobiol 2008;62:1787–93.

[42] Stevens M, Hardman CJ, Stubbins CL. Conspicuousness, not eye mimicry, makes 'eyespots' effective antipredator signals. Behav Ecol 2008;19:525–31.

[43] Stevens M. The role of eyespots as anti-predator mechanisms, principally demonstrated in the Lepidoptera. Biol Rev 2005;80:573–88.

[44] Brakefield PM, Gates J, Keys D, Kesbeke F, Wijngaarden PJ, Monteiro A, et al. Development, plasticity and evolution of butterfly eyespot patterns. Nature 1996;384:236–42.

[45] Gurevitch J. Variation in leaf dissection and leaf energy budgets among populations of Achillea from an altitudinal gradient. Am J Bot 1988;75:1298–306.

[46] Givnish TJ. Comparative studies of leaf form – assessing the relative roles of selective pressures and phylogenetic constraints. New Phytol 1987;106:131–60.

[47] Bharathan G, Goliber TE, Moore C, Kessler S, Pham T, Sinha NR. Homologies in leaf form inferred from KNOXI gene expression during development. Science 2002;296:1858–60.

[48] Hay A, Tsiantis M. The genetic basis for differences in leaf form between Arabidopsis thaliana and its wild relative *Cardamine hirsuta*. Nat Genet 2006;38:942–7.

[49] Johnson AE, Tanner ME. Epimerization via carbon-carbon bond cleavage. L-ribulose-5-phosphate 4-epimerase as a masked class II aldolase. Biochemistry 1998;37:5746–54.

[50] Cheng CC-H. Evolution of the diverse antifreeze proteins. Curr Opin Genet Dev 1998;8:715–20.

[51] Shackleton NJ, Backman J, Zimmerman H, Kent DV, Hall MA, Roberts DG, et al. Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North-Atlantic region. Nature 1984;307:620–3.

[52] Chen LB, DeVries AL, Cheng CHC. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. Proc Natl Acad Sci USA 1997;94:3817–22.

[53] Rothschild LJ. The evolution of photosynthesis… again? Philos Trans Ro Soc B Biol Sci 2008;363:2787–801.

[54] Rodrigues JF, Wagner A. Evolutionary plasticity and innovations in complex metabolic reaction networks. PLOS Comput Biol 2009;5:e1000613.

[55] Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 1999;27:29–34.

[56] Samal A, Rodrigues JFM, Jost J, Martin OC, Wagner A. Genotype networks in metabolic reaction spaces. BMC Syst Biol 2010;4:30.

[57] Rodrigues JF, Wagner A. Genotype networks in sulfur metabolism. BMC Syst Biol 2010;5:39.

[58] Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 2007;3.

[59] Stone J, Wray G. Rapid evolution of *cis*-regulatory sequences via local point mutations. Mol Biol Evol 2001;18:1764–70.

[60] Martchenko M, Levitin A, Hogues H, Nantel A, Whiteway M. Transcriptional rewiring of fungal galactose-metabolism circuitry. Curr Biol 2007;17:1007–13.

[61] Tanay A, Regev A, Shamir R. Conservation and evolvability in regulatory networks: The evolution of ribosomal regulation in yeast. Proc Natl Acad Sci USA 2005;102:7203–8.

[62] Gasch AP, Moses AM, Chiang DY, Fraser HB, Berardini M, Eisen MB. Conservation and evolution of *cis*-regulatory systems in ascomycete fungi. PLOS Biol 2004;2:2202−19.

[63] Wray G, Hahn M, Abouheif E, Balhoff J, Pizer M, Rockman M, et al. The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol 2003;20:1377−419.

[64] Ludwig MZ, Bergman C, Patel NH, Kreitman M. Evidence for stabilizing selection in a eukaryotic enhancer element. Nature 2000;403:564−7.

[65] Maduro M, Pilgrim D. Conservation of function and expression of unc-119 from two *Caenorhabditis* species despite divergence of non-coding DNA. Gene 1996;183:77−85.

[66] Romano L, Wray G. Conservation of Endo16 expression in sea urchins despite evolutionary divergence in both *cis* and *trans*-acting components of transcriptional regulation. Development 2003;130:4187−99.

[67] Jaeger J, Surkova S, Blagov M, Janssens H, Kosman D, Kozlov K, et al. Dynamic control of positional information in the early *Drosophila* embryo. Nature 2004;430:368−71.

[68] Sanchez L, Chaouiya C, Thieffry D. Segmenting the fly embryo: logical analysis of the role of the segment polarity cross-regulatory module. Int J Dev Biol 2008;52:1059−75.

[69] Albert R, Othmer HG. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. Journal of Theoretical Biology 2003;223:1−18.

[70] Ingolia NT. Topology and robustness in the *Drosophila* segment polarity network. PLOS Biol 2004;2:805−15.

[71] MacCarthy T, Seymour R, Pomiankowski A. The evolutionary potential of the *Drosophila* sex determination gene network. J Theor Biol 2003;225:461−8.

[72] Mjolsness E, Sharp DH, Reinitz J. A connectionist model of development. J Theor Biol 1991;152:429−53.

[73] Reinitz J, Mjolsness E, Sharp DH. Model for cooperative control of positional information in *Drosophila* by bicoid and maternal hunchback. J Exp Zool 1995;271:47−56.

[74] Reinitz J. Gene circuits for eve stripes: reverse engineering the Drosophila segmentation gene network. Biophys J 1999;76: A272−A272.

[75] Sharp DH, Reinitz J. Prediction of mutant expression patterns using gene circuits. BioSystems 1998;47:79−90.

[76] Azevedo RBR, Lohaus R, Srinivasan S, Dang KK, Burch CL. Sexual reproduction selects for robustness and negative epistasis in artificial gene networks. Nature 2006;440:87−90.

[77] Bornholdt S, Sneppen K. Robustness as an evolutionary principle. Proc R Soc Lon B Biol Sci 2000;267:2281−6.

[78] Wagner A. Does evolutionary plasticity evolve? Evolution 1996;50:1008−23.

[79] Siegal M, Bergman A. Waddington's canalization revisited: developmental stability and evolution. Proc Natl Acad Sci USA 2000;99:10528−10532

[80] Bergman A, Siegal M. Evolutionary capacitance as a general feature of complex gene networks. Nature 2003;424:549−52.

[81] Nochomovitz YD, Li H. Highly designable phenotypes and mutational buffers emerge from a systematic mapping between network topology and dynamic output. Proc Natl Acad Sci USA 2006;103:4180−5.

[82] Raman K, Wagner A. Evolvability and robustness in a complex signaling circuit. Mol BioSyst 2011;7:1081−92.

[83] Ciliberti S, Martin OC, Wagner A. Innovation and robustness in complex regulatory gene networks. Proc Natl Acad Sci USA 2007;104:13591−13596

[84] Ciliberti S, Martin OC, Wagner A. Circuit topology and the evolution of robustness in complex regulatory gene networks. PLOS Comput Biol 2007;3(2):e15.

[85] Maynard-Smith J. Natural selection and the concept of a protein space. Nature 1970;255:563−4.

[86] Lipman D, Wilbur W. Modeling neutral and selective evolution of protein folding. Proc R Soc Lon B 1991;245:7−11.

[87] Schuster P, Fontana W, Stadler P, Hofacker I. From sequences to shapes and back − a case-study in RNA secondary structures. Proc R Soc Lon B 1994;255:279−84.

[88] Berman H, Battistuz T, Bhat T, Bluhm W, Bourne P, Burkhardt K, et al. The Protein Data Bank. Acta Crystallogr B Biol Crystallogr 2002;58:899−907.

[89] Hofacker I, Fontana W, Stadler P, Bonhoeffer L, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. Monatshefte fuer Chemie 1994;125:167−88.

[90] Flamm C, Fontana W, Hofacker I, Schuster P. RNA folding at elementary step resolution. RNA 2000;6:325−38.

[91] Fontana W, Schuster P. Shaping space: the possible and the attainable in RNA genotype−phenotype mapping. J Theor Biol 1998;194:491−515.

[92] Sumedha, Martin OC, Wagner A. New structural variation in evolutionary searches of RNA neutral networks. BioSystems 2007;90:475−85.

[93] Wagner A. Robustness and evolvability: a paradox resolved. Proc R Soc Lon B Biol Sci 2008;275:91−100.

[94] Jackson R, Kaminski A. Internal initiation of translation in eukaryotes: The picornavirus paradigm and beyond. RNA 1995;1:985−1000.

[95] Mandl C, Holzmann H, Meixner T, Rauscher S, Stadler P, Allison S, et al. Spontaneous and engineered deletions in the 3′ noncoding region of tick-borne encephalitis virus: construction of highly attenuated mutants of a flavivirus. J Virol 1998;72:2132−40.

[96] Iserentant D, Fiers W. Secondary structure of messenger-RNA and efficiency of translation initiation. Gene 1980;9:1−12.

[97] Feist AM, Palsson BO. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat Biotech 2008;26:659−67.

[98] Cotterell J, Sharpe J. An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. Mol Syst Biol 2010;6:425.

[99] Doolittle R. The origins and evolution of eukaryotic proteins. Philos Trans Ro Soc B Biol Sci 1995;349:235−40.

[100] Thornton J, Orengo C, Todd A, Pearl F. Protein folds, functions and evolution. J Mol Biol 1999;293:333−42.

[101] Todd A, Orengo C, Thornton J. Evolution of protein function, from a structural perspective. Curr Opin Chem Biol 1999;3:548−56.

[102] Bastolla U, Porto M, Roman HE, Vendruscolo M. Connectivity of neutral networks, overdispersion, and structural conservation in protein evolution. J Mol Evol 2003;56:243−54.

[103] Rost B. Enzyme function less conserved than anticipated. J Mol Biol 2002;318:595–608.

[104] Aronson H, Royer W, Hendrickson W. Quantification of tertiary structural conservation despite primary sequence drift in the globin fold. Prot Sci 1994;3:1706–11.

[105] Hardison RC. A brief history of hemoglobins: Plant, animal, protist, and bacteria. Proc Natl Acad Sci USA 1996;93:5675–9.

[106] Goodman M, Pedwaydon J, Czelusniak J, Suzuki T, Gotoh T, Moens L, et al. An evolutionary tree for invertebrate globin sequences. J Mol Evol 1988;27:236–49.

[107] Copley RR, Bork P. Homology among $(\beta\alpha)_8$ barrels: Implications for the evolution of metabolic pathways. J Mol Biol 2000;303:627–40.

[108] Wierenga RK. The TIM-barrel fold: a versatile framework for efficient enzymes. FEBS Lett 2001;492:193–8.

[109] Schuster P. Molecular insights into evolution of phenotypes. In: Crutchfield JP, Schuster P, editors. Evolutionary dynamics: Exploring the Interplay of Selection, Accident, Neutrality, and Function. New York, NY: Oxford University Press; 2003. p. 163–215.

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

[111] Huynen MA. Exploring phenotype space through neutral evolution. J Mol Evol 1996;43:165–9.

[112] Eigen M. Viral Quasi-species. Scientific American 1993;269:42–9.

[113] True JR, Carroll SB. Gene co-option in physiological and morphological evolution. Annu Rev Cell Dev Biol 2002;18:53–80.

[114] Panganiban G, Rubenstein JLR. Developmental functions of the Distal-less/Dlx homeobox genes. Development 2002;129:4371–86.

[115] Huang W, Petrosino J, Hirsch M, Shenkin P, Palzkill T. Amino acid sequence determinants of beta-lactamase structure and activity. J Mol Biol 1996;258:688–703.

[116] Rennell D, Bouvier S, Hardy L, Poteete A. Systematic mutation of bacteriophage T4 lysozyme. J Mol Biol 1991;222:67–87.

[117] Weatherall DJ, Clegg JB. Molecular genetics of human haemoglobin. Annu Rev Genet 1976;10:157–78.

[118] Kleina L, Miller J. Genetic studies of the lac repressor. 13. Extensive amino-acid replacements generated by the use of natural and synthetic nonsense suppressors. J Mol Biol 1990;212:295–318.

[119] Wang Z, Zhang J. Abundant indispensable redundancies in cellular metabolic networks. Genome Biol Evol 2009;1:23–33.

[120] Blank LM, Kuepfer L, Sauer U. Large-scale C-13-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. Genome Biol 2005;6:R49.

[121] Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED. Metabolic network structure determines key aspects of functionality and regulation. Nature 2002;420:190–3.

[122] Segre D, Vitkup D, Church G. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 2002;99:15112–7

[123] Edwards JS, Palsson BO. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci USA 2000;97:5528–33.

[124] Bornberg-Bauer E, Chan H. Modeling evolutionary landscapes: mutational stability, topology, and superfunnels in sequence space. Proc Natl Acad Sci USA 1999;96:10689–94

[125] Meiklejohn C, Hartl D. A single mode of canalization. Trends Ecol Evol 2002;17:468–73.

[126] Wagner A. Robustness and Evolvability in Living Systems. Princeton, NJ: Princeton University Press; 2005.

[127] Wagner GP, Booth G, Bagherichaichian H. A population genetic theory of canalization. Evolution 1997;51:329–47.

[128] Papp B, Teusink B, Notebaart RA. A critical view of metabolic network adaptations. HFSP J 2009;3:24–35.

[129] Soyer OS, Pfeiffer T. Evolution under fluctuating environments explains observed robustness in metabolic networks. PLOS Comput Biol 2010;6:e1000907.

[130] Vitkup D, Kharchenko P, Wagner A. Influence of metabolic network structure and function on enzyme evolution. Genome Biol 2006;7:R39.

[131] Papp B, Pal C, Hurst LD. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 2004;429:661–4.

[132] Nishikawa T, Gulbahce N, Motter A. E. Spontaneous Reaction Silencing in Metabolic Optimization. PLOS Comput Biol 2008;4:e1000236

[133] Freilich S, Kreimer A, Borenstein E, Gophna U, Sharan R, Ruppin E. Decoupling environment-dependent and independent genetic robustness across bacterial species. PLOS Comput Biol 2010;6:e1000690.

[134] Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol 2003;4:R54.

[135] Thomas GH, Zucker J, MacDonald SJ, Sorokin A, Goryanin I, Douglas AE. A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. BMC Syst Biol 2009;3:24.

[136] Pal C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD. Chance and necessity in the evolution of minimal metabolic networks. Nature 2006;440:667–70.

[137] Yus E, Maier T, Michalodimitrakis K, van Noort V, Yamada T, Chen WH, et al. Impact of genome reduction on bacterial metabolism and its regulation. Science 2009;326:1263–8.

[138] Benfey PN, Mitchell-Olds T. Perspective — from genotype to phenotype: systems biology meets natural variation. Science 2008;320:495–7.