Review

Genotype networks shed light on evolutionary constraints

Andreas Wagner

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

An evolutionary constraint is a bias or limitation in phenotypic variation that a biological system produces. One can distinguish physicochemical, selective, genetic and developmental causes of such constraints. Here, I discuss these causes in three classes of system that bring forth many phenotypic traits and evolutionary innovations: regulatory circuits, macromolecules and metabolic networks. In these systems, genotypes with the same phenotype form large genotype networks that extend throughout a vast genotype space. Such genotype networks can help unify different causes of evolutionary constraints. They can show that these causes ultimately emerge from the process of development; that is, how phenotypes form from genotypes. Furthermore, they can explain important consequences of constraints, such as punctuated stasis and canalization.

Evolutionary constraints

No organism or species can produce every conceivable kind of phenotypic variation. This limitation is encapsulated in the notion of constraints on phenotypic evolution [1]. An evolutionary constraint is a bias or limitation in phenotypic variation that a biological system produces. Extreme examples include the absence of photosynthesis in higher animals, the general lack of teeth in the lower jaw of frogs, the absence of palm trees in cold climates and the absence of birds that give birth to live young instead of to eggs [1,2]. In these examples, a trait is completely absent. More subtle constraints manifest themselves as correlations among different characters. A paradigmatic case is allometric scaling [2]. Here, the value of one quantitative trait constrains that of another, typically via a specific nonlinear relationship, For example, metabolic rate is proportional to body mass m raised to approximately three-quarter power ($m^{0.75}$) across many different species [3].

It is not hard to see that constraints can influence the spectrum of evolutionary adaptations and innovations that are accessible to living things. For this reason, questions about the causes and consequences of constrained evolution have attracted much attention [1,4–10]. There are multiple kinds and causes of phenotypic constraint (Box 1), which are often difficult to disentangle. Part of the problem is that there is poor understanding of how genotypic change translates into phenotypic change, at least for complex, macroscopic traits.

Constraints and the genotype–phenotype relationship

In recent years, several genetic system classes whose genotype–phenotype relationships can be analyzed more easily than those of macroscopic phenotypes have become accessible to investigation. These system classes are metabolic networks, gene regulatory circuits and macromolecules (i.e. protein and RNA). They are also of interest in their own right, because they are involved in most, if not all, macroscopic phenotypic characters and, thus, also in evolutionary innovations. The genotype–phenotype relationships of these systems share several important features that facilitate such innovations [11–20]. For these reasons, these systems are attractive subjects to study evolutionary constraints. A better understanding of their genotype–phenotype relationships could lead to improved understanding of the causes of constraints.

In this Review, I highlight how these systems can help understand constrained phenotypic evolution. I discuss the entangled causes of phenotypic constraints, and then summarize what recent work has revealed about the genotype–phenotype relationships of metabolic networks, gene regulatory circuits and macromolecules. I show that these relationships can accommodate all four causes of constrained evolution. Most importantly, I argue that these causes can all be viewed as consequences of a single process: how phenotypes form from information contained in genotypes, or the analog of 'development' for these systems.

Entangled causes of phenotypic constraints

Box 1 highlights four major causes of phenotypic constraint: physicochemical, selective, genetic and developmental. They are not mutually exclusive, and can overlap in ways that can be difficult to disentangle, especially for complex morphological traits. A case in point is constrained variation in segment number and identity in the fruit fly *Drosophila melanogaster*. During the 1980s, researchers screened thousands of fly mutants created in large-scale mutagenesis experiments. These screens revealed only a small number of variants in segment number, orientation and identity [21]. Among these variants were embryos that lacked several consecutive segments and embryos that lacked every other (odd-numbered or even-numbered) segment. At first, genetic constraints might seem the best candidate cause for limited variation in such a genetic screen. However, because a complex developmental process is involved in segmentation, development itself can be the cause of this constrained variation.

Corresponding author: Wagner, A. (andreas.wagner@ieu.uzh.ch).

0169-5347/$ – see front matter doi:10.1016/j.tree.2011.07.001 Trends in Ecology and Evolution, November 2011, Vol. 26, No. 11

Box 1. Four major, but not mutually exclusive causes of evolutionary constraints

Physicochemical constraints are the first major cause of evolutionary constraints on phenotypic variation. For example, all organisms above a given (small) size must have a circulatory system, because diffusion is not fast enough to deliver nutrients to all body parts of a large organism.

The second major cause is selective constraints, which result from the action of natural selection. An extreme example involves cyclopia, a condition in which only one (central) eye forms during the development of an animal. In zebrafish, cyclopic mutants can be created in mutagenesis screens, but are lethal [73] and would immediately be eliminated by natural selection. More subtle selective constraints are everywhere, because natural selection affects most phenotypes. For example, most mutations in the coding region of a protein have deleterious albeit often subtle fitness effects [74,75], and would be eliminated by natural selection over time. By influencing the frequency of such variants, selection constrains the distribution of protein phenotypes within a population.

Third, in a genetic constraint, any one genotype or its mutants can only produce a small subset (or none) of a broad spectrum of conceivable phenotypic variants. Genetic constraints have been known for a long time. A candidate example involves variation in wing shapes and eye morphology that occurs readily through mutations in the fly *Drosophila subobscura*, but not in its relative *Drosophila melanogaster* [1,76].

A fourth cause is developmental constraints, which emerge from the processes that produce phenotypes from genotypes [1]. A classic example regards variation in the number of digits in salamanders and frogs [7,22,77]. The salamander *Ambystoma mexicanum* (the axolotl) has a hindlimb with five toes or digits that are conventionally labeled I–V. During development, digit I forms first and digit V forms last. In salamanders related to the axolotl, one or more of these digits is lost, and the lost digits are those that form latest in development. Digits get lost in a specific order, because of how they originate in development, namely through groups of cells that produce cartilage where the digits will later form. The number of such cartilaginous cell groups reflects the number of digits [7,22,77]. In sum, evolutionary variation in a trait (order of digit loss) is constrained here by trait development. Because development involves many different processes, including tissue movements, hormone action and wave formation in excitable media [10,59,72], developmental constraints can themselves have different origins [10].

To make matters more complicated, fly segmentation requires several, now well-characterized genes, which entangles the developmental and genetic causes of constraints. In addition, segmentation involves diffusion of molecules along an embryo and chemical interactions between gene products, making it potentially subject to physicochemical constraints [22]. Finally, past selection that favored and stabilized constant segment numbers and identities might also have contributed to such constrained variation. After all, other arthropods have more variable segment numbers than *Drosophila*. In sum, different causes of constraints are entangled in this example, and such entanglement is the rule rather than the exception.

Why study constraints for metabolic, regulatory and molecular phenotypes?

All complex macroscopic traits comprise microscopic, submicroscopic and molecular traits, down to the level of DNA. Likewise, DNA change can percolate all the way up to macroscopic traits. Although an understanding of the full complexity of this hierarchical organization is beyond current means, important systems can be studied that are necessary to form complex traits and changes therein. Here, I introduce three classes of such system. They are crucial to understanding phenotypic constraints, because they are involved in forming many phenotypes.

Genome-scale metabolic networks

The first class of system comprises genome-scale metabolic networks. These are networks of hundreds to thousands of chemical reactions (catalyzed by enzymes that are encoded by genes) that synthesize all small molecules in biomass from environmental nutrients. In addition, they produce energy and many important secondary metabolites. The metabolic phenotypes of such networks are involved in many traits and in many metabolic innovations, from within microbes to higher organisms. Examples include the ability of microbes to grow on synthetic antibiotics or other toxic xenobiotic compounds [23–27], and the urea cycle of land-living animals [28]. Such metabolic innovations often involve new combinations of chemical reactions (i.e. enzymes), frequently mediated by horizontal gene transfer, that already exist elsewhere [26,28].

Regulatory circuits

The second system class comprises regulatory circuits. These involve interacting gene products that influence the biological activity of each other. 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 [22,29]. 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 circuit are transcriptional regulation circuits, because transcriptional regulation provides a backbone of regulation to most organisms, and because such circuits drive many pattern formation processes during embryonic development. Among the best-known examples are Hox genes involved in patterning limbs and many other body structures in animals, as well as MADS box genes involved in patterning flowers [30–35]. Regulatory change is also often involved when new macroscopic phenotypes form, such as the dissected leaves of some flowering plants or the eyespots of some butterflies [29,36].

Macromolecules

The final system class comprises protein and RNA macromolecules. Albeit at the lowest rung of biological organization, they serve many important functions: not only do they catalyze thousands of chemical reactions, but they also exchange many chemicals between cells and their environment, give structural support to cells, and are central to locomotion and transport. Not surprisingly, important phenotypic adaptations are directly traceable to changes in macromolecules. Examples include antifreeze proteins that can enable an organism to survive at low temperatures, and globin molecules adapted to oxygen transport at altitudes exceeding 10 kilometers [37–40].

In all three system classes, it has recently become possible to gain a better understanding of the relationship between genotype and phenotype. One reason is that sufficiently large amounts of empirical data exist to analyze this relationship, as in the case of proteins. Another reason is that the ability to model biological systems quantitatively has improved [4–12,12–15,15,16,16–47]. Here, I summarize insights from recent work on these system classes [11–20].

Genotype spaces and their organization

Genotypes in all three system classes exist in a vast genotype space. Although such genotypes are ultimately DNA sequences, it is often useful to represent them in more compact ways. For example, the genotype of a metabolic network can be characterized through the presence or absence of enzyme-catalyzed reactions in the network. The genotype space of metabolic networks is the space of all possible metabolic networks. The current known 'universe' of metabolic reactions comprises more than 5000 such reactions, each of which could be present or absent in any one network. Thus, metabolic genotype space is a vast hyperastronomical space comprising more than 2^{5000} metabolic networks [11,12].

The genotypes of regulatory circuits are usefully represented as patterns of regulatory interactions. Such regulatory interactions can be enormously complex, involving cooperative interactions among multiple regulators. However, models of specific kinds of regulatory circuit can make accurate predictions under sensible simplifying assumptions, such as assuming that most regulatory interactions involve pairs of molecular species [41]. In a transcriptional regulation circuit, for example, any one gene X can have an activating, repressing, or no effect on the expression of another gene Y, as determined by regulatory DNA sequences near gene Y. The space of all circuits then comprises all possible pairwise patterns of interactions between a given set of genes [14–16,48,49]. These interactions give rise to the gene expression phenotype of a circuit.

The genotypes of molecules also exist in a space of amino acid and nucleotide sequences. Characterization of this genotype space has the longest research history [17,19,50].

The genotype spaces of these three system classes share several similarities [11–15,17–20]. The first pertains to the neighbors of a genotype. A neighbor of a genotype differs in one system part (e.g. a chemical reaction, a regulatory interaction, or an amino acid) from the genotype itself. The mutations that distinguish two neighbors include DNA deletions that impair an enzyme-coding gene (metabolic networks), mutations of regulatory DNA that affect a single regulatory interaction (regulatory circuits) and amino acid or nucleotide point mutations (macromolecules). In all three system classes, individual genotypes typically have many neighbors with the same phenotype. Wherever many neighboring genotypes have the same phenotype, genotypes are, to some extent, robust to mutations. That is, their phenotype does not change in response to some mutations.

Second, genotypes with the same phenotype form vast connected genotype networks that reach throughout genotype space. This means that one can step through a series of small genetic changes from one genotype to its neighbor, to the neighbor's neighbor, and so on, without ever changing a phenotype. Very different genotypes can thus have the same phenotype.

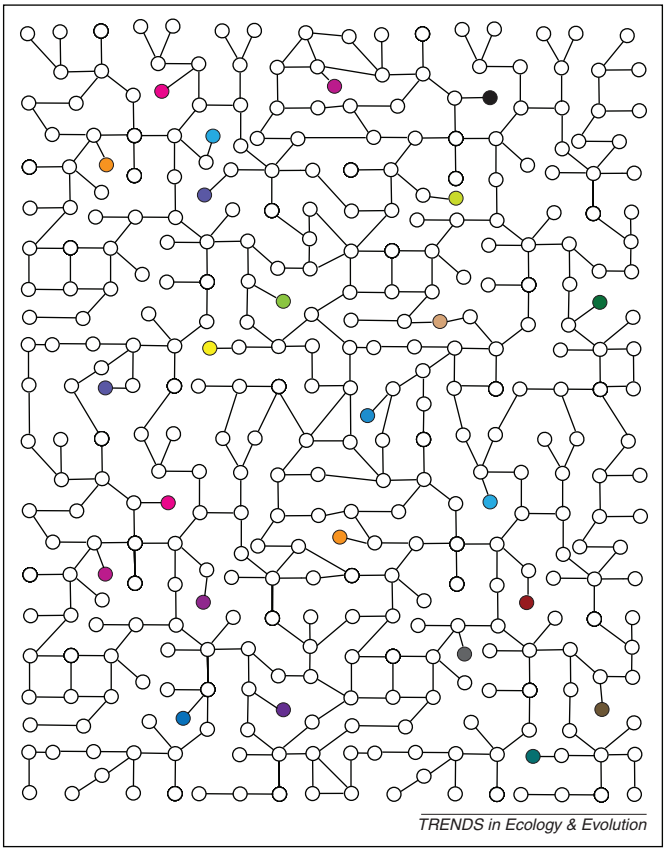

Figure 1. Genotype networks in genotype space. The figure schematically represents a hypothetical set of genotypes (open circles) in genotype space (large rectangle) that share the same phenotype and that form a genotype network. Neighboring genotypes are connected by straight lines. Colored circles indicate genotypes with different phenotypes (each color corresponds to a different phenotype) that are neighbors of genotypes on the genotype network. The figure shows that many different novel phenotypes can be accessed from a connected genotype network that extends throughout genotype space, and that different regions of genotype space can contain different novel phenotypes. Note that any two-dimensional illustration of a high-dimensional genotype space has severe limitations; for example, each of the colored genotypes would form part of a large genotype network that the figure does not display, and individual genotypes can have thousands of neighbors, many more than the figure can show [53].

Third, based on what is currently known, the neighborhoods of any two genotypes on the same genotype network (i.e. with the same phenotype) contain very different novel phenotypes (Figure 1). This assertion is based on evidence from proteins [20], RNA [17,51,52], model regulatory circuits [15] and metabolic networks [11,13]. Together, these properties facilitate the evolution of novel phenotypes through exploration of genotype space. They enable a population to keep its phenotype unchanged while exploring different regions of genotype space and the many novel phenotypes therein [53].

The three system classes I discuss here are simpler and more tractable than those giving rise to macroscopic phenotypes. However, that is not the only reason to use them to study constrained phenotypic evolution. A second reason is that one can compare the phenotypes in any one of these systems quantitatively and analogously to macroscopic traits in a morphospace [54,55]. A third reason is that,

in all three system classes, phenotypes are formed by complex processes that can themselves constrain variation, yet they avoid the still intractable complexity of embryonic development. For protein phenotypes, the relevant process is protein folding; for metabolic phenotypes, it is the flow of metabolites through a reaction network; and, for gene activity phenotypes, it is the dynamical change of gene activities caused by regulatory interactions. The latter process can capture important aspects of the dynamical complexities of developmental pattern formation, such as static geometric patterns and traveling waves in the activities of regulatory molecules [41,42,56–59]. How this process produces phenotypes is therefore relevant for the understanding of macroscopic traits.

I next revisit the four causes for phenotypic constraints in the context of these three system classes, highlighting an important aspect of their relationship; that is, that the processes of phenotype production are the root cause of other constraints.

Physicochemical constraints

Physicochemical factors constrain observable protein structure phenotypes. To give but one example, consider the packing of hydrophobic amino acids in the core of globular proteins, an early, important event in protein folding. As a result, polar -NH and -CO groups in the backbone of buried amino acids cannot form energetically favorable hydrogen bonds with water. Proteins circumvent this problem by forming α-helix and β-sheet secondary structure elements, where amino acids form these hydrogen bonds with each other [60]. Inappropriate exposure of hydrophobic amino acids disrupts protein function and can lead to insoluble proteins owing to aggregation. The packing requirement of hydrophobic amino acids is part of the reason why only a small fraction of random amino acid sequences fold, and why the total number of protein structure phenotypes is modest [61]. In other words, this physicochemical requirement constrains the set of allowable protein phenotypes.

The phenotypes of metabolic networks are also subject to physicochemical constraints. These are primarily dictated by organic chemistry; that is, by the organic chemical reactions that can occur in water. Although little is known about the set of such allowable reactions, examples of physicochemical constraints on metabolism are easy to come by. An especially simple one relates to the ability to build biomass from a source of carbon or other chemical elements. To do this, the metabolic network of an organism needs to have a minimal number of chemical reactions [12,13]. In other words, metabolic phenotypes that allow growth on a variety of sources are constrained by metabolic network size.

Physicochemical factors that can constrain the gene activity phenotypes of regulatory circuits are still poorly understood. On the one hand, regulatory interactions are often mediated by very short amino acid and DNA sequences, for example in transcriptional regulation or protein phosphorylation [62–66]. Such short sequences can evolve rapidly and might endow regulatory circuits with flexible, almost unconstrained phenotypes. On the other hand, factors such as ubiquitous molecular noise inside a cell might constrain attainable patterns of gene expression and molecular activities [67–69].

The processes that create phenotypes from genotypes for the systems I examine here are protein folding, the dynamically changing regulatory interactions within regulatory circuits and the chemical synthesis of biomass molecules. These processes are the analog of 'development' for these systems. Taken together, the above observations show that these processes are key to understanding physicochemical constraints.

Selective constraints

Genotype spaces and genotype networks also provide a systematic framework to understand selective constraints. Such constraints are as ubiquitous for the phenotypes I consider here as they are for macroscopic traits.

An organism with a mutated metabolic network, where, say, an enzyme-coding gene has suffered a loss of function mutation or a mutation that reduces the catalytic efficiency of an enzyme, might produce biomass at a lower rate and, thus, slow down cell growth. As a result, natural selection might eliminate organisms hosting such mutations over time. In this way, selection can constrain the distribution of protein or metabolic network phenotypes.

Selective constraints also influence the phenotypes of regulatory circuits. The experimental literature, especially in cell and developmental biology, is full of mutations that change the gene activity phenotype of a regulatory circuit, such that an organismal phenotype (be it that of a cell or a multicellular organism) does not form properly. Some such mutations can only be seen after mutagenesis in the laboratory and might rarely, if ever, occur in the wild. However, because experiments can reveal such variants, they are neither prohibited by genetic nor by developmental constraints, but by their detrimental effects on the organism. Their absence in the wild is therefore a result of selective constraints.

Constraints such as these are usefully viewed through the lens of genotype networks, where selection can confine a population to a given genotype network. This perspective explains how the genotype of a system subject to selective constraints might change substantially while its phenotype remains unchanged.

Aside from these obvious selective constraints, there also exists a less obvious kind of selective constraint. It emerges from ongoing selection favoring the preservation of existing phenotypes. In the language of population genetics, this kind of selection is called stabilizing selection. Among students of development, it is also known as canalizing selection [1]. Canalizing selection can reduce phenotypic variability, and the genotype space framework can help explain why (Box 2, Figure I). Briefly, phenotypic variability can decrease because genotype networks are internally heterogeneous, containing regions where individual genotypes have more neighbors with the same phenotype than elsewhere.

In addition, the genotype network framework can help understand how the kind of selective constraints described here are entangled with the processes producing phenotypes from genotypes. It is these processes that are fundamentally responsible for how genotypes map onto

Box 2. Canalizing selection

Canalizing selection disfavors phenotypes that deviate from a given, optimal phenotype, whether they arise through rare genetic perturbations or through more frequent non-genetic perturbations, such as gene expression noise [68,69,78,79]. Sustained canalizing selection can increase the robustness of an evolving population to genetic and non-genetic perturbations [80,81]. The selected phenotype becomes less variable (more constrained) in response to these perturbations. Candidate examples from macroscopic to molecular phenotypes have been known for some time [82–84]. In wild populations subject to canalizing selection, more recently evolved traits can show greater variability compared with older traits, because canalizing selection has acted on the more recent traits for a shorter amount of time. This holds, for instance, for rows of bristles used in the male courtship of *Drosophila sylvestris* fruit flies. Here, newly evolved rows of bristles can be more variable than older rows [85]. Among molecular phenotypes, RNA secondary structures provide well-studied examples [86–89]. Eukaryotic miRNA precursor molecules, parts of RNA viral genomes, and single-stranded RNA pathogens called viroids, whose secondary structure phenotypes are under stabilizing selection, have become phenotypically less variable in response to perturbations that include mutations and thermal noise [86–89].

The genotype space framework provides a simple explanation for how canalizing selection can reduce phenotypic variability, at least where selection favors genotypes robust to mutations (Figure I). The explanation is that genotype networks are internally heterogeneous objects. In some regions of any one genotype network, genotypes have more neighbors with the same phenotype, whereas in others they have fewer neighbors with the same phenotype. When subject to frequent perturbations, a population will tend to accumulate in a region of a genotype network where most genotypes have many neighbors with the same phenotypes; that is, where genotypes are highly robust to mutation. The reason is that, in such regions, fewer perturbations cause detrimental phenotypic change, and a greater number of perturbed individuals survive [80,81,90,91]. In such regions, individuals also produce little or no phenotypic variation. In other words, once a population enters such a region, its phenotype becomes highly constrained.

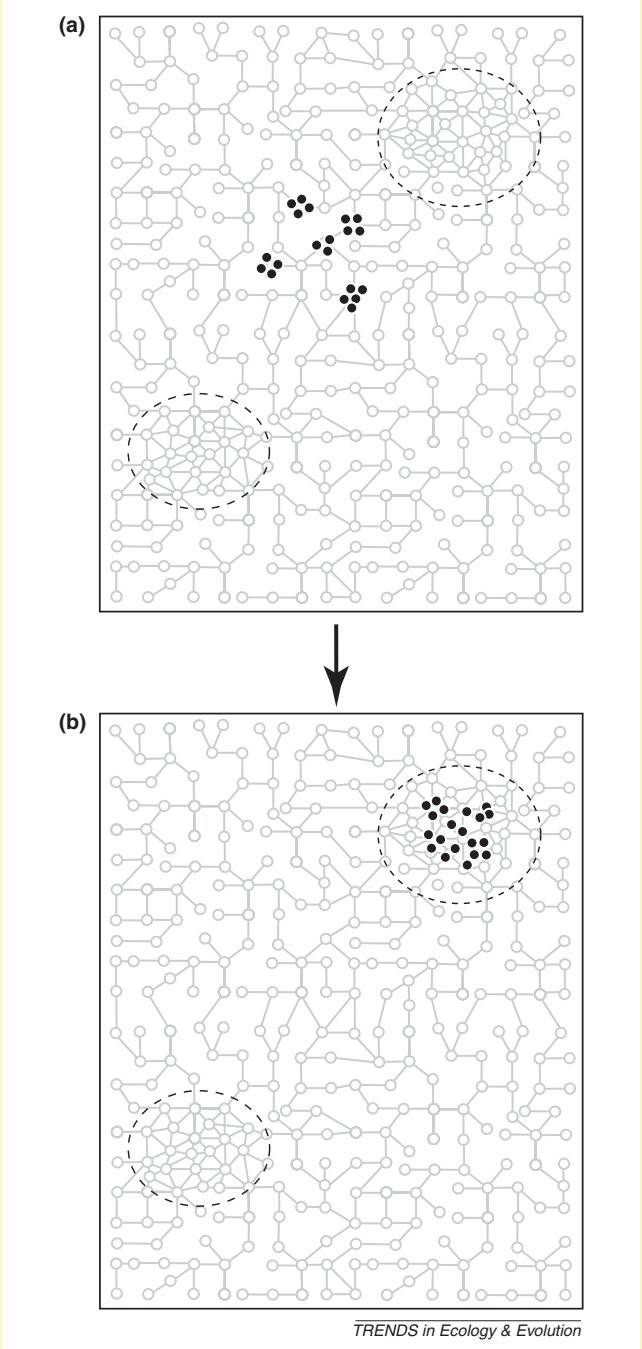

Figure I. The increase in phenotypic constraints by canalizing selection owing to internal genotype network heterogeneity. Open circles connected by straight lines represent neighboring genotypes with the same phenotype on a hypothetical genotype network. Broken ellipses correspond to regions of the genotype network where individuals have more neighbors with the same phenotype and, thus, higher robustness. **(a)** shows a population of genotypes (black circles) on the genotype network. Under selection that favors maintenance of the phenotype, the population can move into regions of high robustness **(b)**. There can be more than one such region, as indicated by the two ellipses. Which of them a population comes to occupy might depend on a variety of factors, such as the initial distance of a population from each region. Analogous considerations hold for robustness to other perturbations, which is often correlated with robustness to mutations. From [53]. Used by permission of Oxford University Press.

phenotypes. They are thus also responsible for the organization of genotype networks, including their internal heterogeneity, without which canalizing selection could not reduce variability. The relationship between genotype and phenotype precedes the action of natural selection on any one population. It determines what natural selection can achieve.

Genetic constraints on phenotypic variation

Genotype space is astronomically vast and contains many phenotypes. This observation holds for all three classes of systems examined; that is, for molecules, regulatory circuits and metabolic networks. The immediate neighborhood of any one genotype contains only a tiny fraction of all possible genotypes. For example, for proteins of 100 amino acids, where genotype space comprises more than 10^{130} amino acid sequences, any one protein genotype G has only $19 \times 100 = 1900$ mutant neighbors, fewer than one $10^{-126\text{th}}$ of genotype space. Any one neighborhood can contain only a tiny fraction of possible phenotypes, whose number can be astronomical for some systems [11,14]. In addition, because genotypes are typically robust to mutation, many of the genotypes in a neighborhood of any genotype G have the same phenotype as G. Taken together, this means that

Box 3. Genotype networks and evolutionary stasis

An absence of evolutionary change, known as stasis, can occur for at least two reasons. First, an organism's phenotype might be optimal in a given environment, and no superior phenotype might exist. Second, phenotypic variability might be present, but the right kind of phenotypic variation (novel adaptive phenotypes) might be absent. The latter kind of stasis often arises when a superior phenotype exists, but when this phenotype has not yet been 'discovered' by an evolving population. A characteristic pattern of evolution in this case is episodic change or punctuated evolution. Here, long periods of evolutionary stasis, where the phenotype of a population changes little, are punctuated by rapid evolutionary change, where a population 'discovers' novel, superior phenotypes. Such episodic change has been observed at all levels of biological organization and on different timescales [18,92–96]. For example, it occurs for: (i) morphological traits observable in the fossil record, where its causes have led to much debate [2]; (ii) cellular traits in laboratory evolution experiments, such as bacterial cell size [95]; and (iii) molecules evolving under directional selection [18,92,96].

The genotype space framework can readily explain such punctuated stasis. It is a period of time where populations of organisms explore an existing genotype network until one mutant 'finds' the genotype network of a superior phenotype. As a population discovers successively better phenotypes, further improvement becomes increasingly difficult, such that the periods of stasis (i.e. genotype network exploration) become longer and longer. For some systems, such as RNA molecules in the search of superior secondary structures, this process has been characterized in great detail [18,96,97]. In sum, this kind of stasis arises because not all phenotypic variation is mutationally accessible from any one genotype or population. In other words, it is a consequence of genetic constraints.

phenotypic variation must be constrained genetically. Any one small neighborhood simply does not contain enough genotypes to harbor all possible phenotypes. The structure of genotype space implies that genetic constraints are inevitable. Beyond this observation, different regions of a genotype space can also preferentially harbor different phenotypes, another source of genetic constraints. This phenomenon has been demonstrated for proteins [20], regulatory circuits [15] and metabolic networks [11–13]. Genetic constraints are also linked to an absence of evolutionary change, known as stasis, an extreme form of constrained phenotypic evolution. Box 3 details how genotype networks can help explain this linkage.

'Developmental' constraints

By themselves, systems in each of the three classes I focus on here cannot produce the macroscopic phenotypes that form during organismal development. However, although these systems are not sufficient for development, each of them is necessary. In addition, their phenotypes form through complex processes, including protein folding and dynamically changing transcriptional regulatory interactions, which unfold in time, similarly to development. Thus, one can examine the role that these 'developmental' processes have in constraining variation. Several pertinent observations emerge from the previous sections. First, the processes of phenotype formation are at the root of physicochemical constraints on phenotypic variation. Second, these processes create internally heterogeneous genotype networks, which are responsible for selective constraints, including those caused by canalizing selection. Third, the distribution of phenotypes in genotype space is a consequence of the processes that produce phenotypes from genotypes. This distribution is the origin of genetic constraints.

Taken together, this means that constraints emerging from the production of phenotypes are ultimately the cause of the three other classes of constraints. This causal role of phenotype production can also help explain why different classes of constraint, such as genetic and developmental constraints, are difficult to disentangle.

Concluding remarks

In the interest of clarity, I used two simplifications here to expose the unifying themes among different evolutionary constraints. The first is that I did not highlight the important role of the environment in forming all phenotypes, from protein structures to macroscopic traits [70,71]. However, relaxing this simplification would not invalidate the genotype space framework and the insights it provides. On the contrary, the environment and its changes help bring genotype networks into existence [53]. Second, I focused on phenotypes that are simpler than the macroscopic traits of higher organisms. What one can learn from them about such traits is limited, for example because they lack some of the complex pattern formation processes that form such traits [10,22,55,59,72]. However, these phenotypes also have tremendous advantages when studying constraints. First, there is a better understanding of the relationship between genotype and phenotype better for them; second, they also form through complex processes analogous to development; and third, they are involved in the process of building macroscopic phenotypes. Students of organismal development and its evolution have long emphasized the need to understand how phenotypes form to understand evolution and phenotypic innovation [1,7,10,29]. The traits I discuss here support this notion, because they show that phenotype formation ultimately gives rise to several other classes of evolutionary constraint. The framework of genotype networks therefore has the potential to clarify and simplify how researchers think of constrained evolution.

Acknowledgements

I thank Joanna Masel, as well as an anonymous reviewer, for valuable comments. I acknowledge support through SNF grants 315200-116814, 315200-119697, and 315230-129708, as well as through the YeastX project of SystemsX.ch. Some material in this article is expanded on in [53] by the same author.

References

1 Maynard-Smith, J. *et al.* (1985) Developmental constraints and evolution. *Q. Rev. Biol.* 60, 265–287
2 Futuyma, D.J. (1998) *Evolutionary Biology*, Sinauer
3 West, G.B. *et al.* (1997) A general model for the origin of allometric scaling laws in biology. *Science* 276, 122–126
4 Brakefield, P.M. (2006) Evo-devo and constraints on selection. *Trends Ecol. Evol.* 21, 362–368
5 Cheverud, J.M. (1984) Quantitative genetics and developmental constraints on evolution by selection. *J. Theor. Biol.* 110, 155–171
6 Hodin, J. (2000) Plasticity and constraints in development and evolution. *J. Exp. Zool.* 288, 1–20
7 Oster, G.F. *et al.* (1988) Evolution and morphogenetic rules – the shape of the vertebrate limb in ontogeny and phylogeny. *Evolution* 42, 862–884
8 Shubin, N.H. and Alberch, P. (1986) A morphogenetic approach to the origin and basic organization of the tetrapod limb. *Evol. Biol.* 20, 319–387

9 Odell, G.M. *et al.* (1981) The mechanical basis of morphogenesis. 1. Epithelial folding and invagination. *Dev. Biol.* 85, 446–462
10 Newman, S.A. and Bhat, R. (2009) Dynamical patterning modules: a 'pattern language' for development and evolution of multicellular form. *Int. J. Dev. Biol.* 53, 693–705
11 Rodrigues, J.F. and Wagner, A. (2009) Evolutionary plasticity and innovations in complex metabolic reaction networks. *PLoS Comput. Biol.* 5, e1000613
12 Samal, A. *et al.* (2010) Genotype networks in metabolic reaction spaces. *BMC Syst. Biol.* 4, 30
13 Rodrigues, J.F. and Wagner, A. (2010) Genotype networks in sulfur metabolism. *BMC Syst. Biol.* 5, 39
14 Ciliberti, S. *et al.* (2007) Circuit topology and the evolution of robustness in complex regulatory gene networks. *PLoS Comput. Biol.* 3, e15
15 Ciliberti, S. *et al.* (2007) Innovation and robustness in complex regulatory gene networks. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13591–13596
16 MacCarthy, T. *et al.* (2003) The evolutionary potential of the *Drosophila* sex determination gene network. *J. Theor. Biol.* 225, 461–468
17 Schuster, P. *et al.* (1994) From sequences to shapes and back – a case-study in RNA secondary structures. *Proc. R. Soc. Lond. B* 255, 279–284
18 Fontana, W. and Schuster, P. (1998) Continuity in evolution: on the nature of transitions. *Science* 280, 1451–1455
19 Lipman, D. and Wilbur, W. (1991) Modeling neutral and selective evolution of protein folding. *Proc. R. Soc. Lond. B* 245, 7–11
20 Ferrada, E. and Wagner, A. (2010) Evolutionary innovation and the organization of protein functions in sequence space. *PLoS ONE* 5, e14172
21 Nusslein-Volhard, C. and Wieschaus, E. (1980) Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795–801
22 Gilbert, S.F. (1997) *Developmental Biology*, Sinauer
23 Rehmann, L. and Daugulis, A.J. (2008) Enhancement of PCB degradation by *Burkholderia xenovorans* LB400 in biphasic systems by manipulating culture conditions. *Biotechnol. Bioeng.* 99, 521–528
24 van der Meer, J.R. *et al.* (1998) Evolution of a pathway for chlorobenzene metabolism leads to natural attenuation in contaminated groundwater. *Appl. Environ. Microbiol.* 64, 4185–4193
25 van der Meer, J.R. (1997) Evolution of novel metabolic pathways for the degradation of chloroaromatic compounds. *Antonie van Leeuwenhoek* 71, 159–178
26 Copley, S.D. (2000) Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. *Trends Biochem. Sci.* 25, 261–265
27 Cline, R.E. *et al.* (1989) Pentachlorophenol measurements in body-fluids of people in log homes and workplaces. *Arch. Environ. Contam. Toxicol.* 18, 475–481
28 Takiguchi, M. *et al.* (1989) Evolutionary aspects of urea cycle enzyme genes. *Bioessays* 10, 163–166
29 Carroll, S.B. *et al.* (2001) *From DNA to Diversity. Molecular Genetics and the Evolution of Animal Design*, Blackwell
30 Hughes, C.L. and Kaufman, T.C. (2002) Hox genes and the evolution of the arthropod body plan. *Evol. Dev.* 4, 459–499
31 Irish, V.F. (2003) The evolution of floral homeotic gene function. *Bioessays* 25, 637–646
32 Wagner, G.P. *et al.* (2003) Hox cluster duplications and the opportunity for evolutionary novelties. *Proc. Natl. Acad. Sci. U.S.A.* 100, 14603–14606
33 Causier, B. *et al.* (2005) Evolution in action: following function in duplicated floral homeotic genes. *Curr. Biol.* 15, 1508–1512
34 Lemons, D. and McGinnis, W. (2006) Genomic evolution of Hox gene clusters. *Science* 313, 1918–1922
35 Hueber, S.D. and Lohmann, I. (2008) Shaping segments: Hox gene function in the genomic age. *Bioessays* 30, 965–979
36 Bharathan, G. *et al.* (2002) Homologies in leaf form inferred from KNOXI gene expression during development. *Science* 296, 1858–1860
37 Cheng, CC-H. (1998) Evolution of the diverse antifreeze proteins. *Curr. Opin. Genet. Dev.* 8, 715–720
38 Chen, L.B. *et al.* (1997) Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc. Natl. Acad. Sci. U.S.A.* 94, 3817–3822
39 Liang, Y.H. *et al.* (2001) The crystal structure of bar-headed goose hemoglobin in deoxy form: the allosteric mechanism of a hemoglobin species with high oxygen affinity. *J. Mol. Biol.* 313, 123–137
40 Golding, G.B. and Dean, A.M. (1998) The structural basis of molecular adaptation. *Mol. Biol. Evol.* 15, 355–369
41 Jaeger, J. *et al.* (2004) Dynamic control of positional information in the early *Drosophila* embryo. *Nature* 430, 368–371
42 Sanchez, L. *et al.* (2008) Segmenting the fly embryo: logical analysis of the role of the Segment Polarity cross-regulatory module. *Int. J. Dev. Biol.* 52, 1059–1075
43 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*. *J. Theor. Biol.* 223, 1–18
44 von Dassow, G. *et al.* (2000) The segment polarity network is a robust development module. *Nature* 406, 188–192
45 Feist, A.M. *et al.* (2009) Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.* 7, 129–143
46 Feist, A.M. and Palsson, B.O. (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* 26, 659–667
47 Reed, J.L. *et al.* (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 4, R54
48 Wagner, A. (2005) Circuit topology and the evolution of robustness in two-gene circadian oscillators. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11775–11780
49 Giurumescu, C.A. *et al.* (2009) Predicting phenotypic diversity and the underlying quantitative molecular transitions. *PLoS Comput. Biol.* 5, e1000354
50 Maynard-Smith, J. (1970) Natural selection and the concept of a protein space. *Nature* 255, 563–564
51 Huynen, M.A. (1996) Exploring phenotype space through neutral evolution. *J. Mol. Evol.* 43, 165–169
52 Sumedha *et al.* (2007) New structural variation in evolutionary searches of RNA neutral networks. *Biosystems* 90, 475–485
53 Wagner, A. (2011) *The Origins of Evolutionary Innovations. A Theory of Transformative Change in Living Systems*, Oxford University Press
54 Gould, S.J. (1991) The disparity of the Burgess shale arthropod fauna and the limits of cladistic-analysis – why we must strive to quantify morphospace. *Paleobiology* 17, 411–423
55 Salazar-Ciudad, I. (2006) Developmental constraints vs. variational properties: how pattern formation can help to understand evolution and development. *J. Exp. Zool. B* 306B, 107–125
56 Uriu, K. *et al.* (2009) Traveling wave formation in vertebrate segmentation. *J. Theor. Biol.* 257, 385–396
57 Pourquie, O. (2003) The segmentation clock: converting embryonic time into spatial pattern. *Science* 301, 328–330
58 Reinitz, J. *et al.* (1995) Model for cooperative control of positional information in Drosophila by bicoid and maternal hunchback. *J. Exp. Zool.* 271, 47–56
59 Kaneko, K. (2011) Characterization of stem cells and cancer cells on the basis of gene expression profile stability, plasticity, and robustness. *Bioessays* 33, 403–413
60 Branden, C. and Tooze, J. (1999) *Introduction to Protein Structure*, Garland
61 Levitt, M. (2009) Nature of the protein universe. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11079–11084
62 Gerhart, J. and Kirschner, M. (1998) *Cells, Embryos, and Evolution*, Blackwell
63 Holt, L.J. *et al.* (2009) Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* 325, 1682–1686
64 Tan, C.S.H. *et al.* (2009) Positive selection of tyrosine loss in metazoan evolution. *Science* 325, 1686–1688
65 Stone, J. and Wray, G. (2001) Rapid evolution of cis-regulatory sequences via local point mutations. *Mol. Biol. Evol.* 18, 1764–1770
66 Wray, G. *et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419
67 Eldar, A. and Elowitz, M.B. (2010) Functional roles for noise in genetic circuits. *Nature* 467, 167–173
68 Raser, J.M. and O'Shea, E.K. (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304, 1811–1814
69 Blake, W. *et al.* (2003) Noise in eukaryotic gene expression. *Nature* 422, 633–637

70 West-Eberhard, M.J. (2003) *Developmental Plasticity and Evolution*, Oxford University Press
71 Dyson, H.J. and Wright, P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208
72 Murray, J.D. (1989) *Mathematical Biology*, Springer
73 Hatta, K. *et al.* (1991) The cyclops mutation blocks specification of the floor plate of the zebrafish central-nervous-system. *Nature* 350, 339–341
74 Sawyer, S.A. *et al.* (2007) Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6504–6510
75 Lynch, M. and Conery, J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155
76 Spurway, H. (1949) Remarks on Vavilov's law of homologous variation. *La Ricerca Scientifica* 19 (Suppl.), 3–9
77 Alberch, P. and Gale, E.A. (1985) A developmental analysis of an evolutionary trend – digital reduction in amphibians. *Evolution* 39, 8–23
78 Bar-Even, A. *et al.* (2006) Noise in protein expression scales with natural protein abundance. *Nat. Genet.* 38, 636–643
79 Pedraza, J.M. and Paulsson, J. (2008) Effects of molecular memory and bursting on fluctuations in gene expression. *Science* 319, 339–343
80 Wagner, A. (2005) *Robustness and Evolvability in Living Systems*, Princeton University Press
81 van Nimwegen, E. *et al.* (1999) Neutral evolution of mutational robustness. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9716–9720
82 Waddington, C.H. (1960) Experiments on canalizing selection. *Genet. Res.* 1, 140–150
83 Smith, J.M. and Sondhi, K.C. (1960) The genetics of a pattern. *Genetics* 45, 1039–1050
84 Rendel, J.M. (1959) Canalization of the scute phenotype of *Drosophila*. *Evolution* 13, 425–439
85 Carson, H.L. and Lande, R. (1984) Inheritance of a secondary sexual character in *Drosophila silvestris*. *Proc. Natl. Acad. Sci. U.S.A.* 81, 6904–6907
86 Sanjuan, R. *et al.* (2006) In silico predicted robustness of viroid RNA secondary structures. I. The effect of single mutations. *Mol. Biol. Evol.* 23, 1427–1436
87 Wagner, A. and Stadler, P. (1999) Viral RNA and evolved mutational robustness. *J. Exp. Zool. B: Mol. Dev. Evol.* 285, 119–127
88 Borenstein, E. and Ruppin, E. (2006) Direct evolution of genetic robustness in microRNA. *Proc. Natl. Acad. Sci. U.S.A.* 103, 6593–6598
89 Szollosi, G.J. and Derenyi, I. (2009) Congruent evolution of genetic and environmental robustness in micro-RNA. *Mol. Biol. Evol.* 26, 867–874
90 Wagner, A. (1996) Does evolutionary plasticity evolve? *Evolution* 50, 1008–1023
91 Ancel, L.W. and Fontana, W. (2000) Plasticity, evolvability, and modularity in RNA. *J. Exp. Zool. B* 288, 242–283
92 Studer, R.A. and Robinson-Rechavi, M. (2009) Evidence for an episodic model of protein sequence evolution. *Biochem. Soc. Trans.* 37, 783–786
93 Adams, K.L. *et al.* (2002) Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9905–9912
94 Knoll, A.H. (1992) The early evolution of eukaryotes – a geological perspective. *Science* 256, 622–627
95 Elena, S.F. *et al.* (1996) Punctuated evolution caused by selection of rare beneficial mutations. *Science* 272, 1802–1804
96 Fontana, W. and Schuster, P. (1998) Shaping space: the possible and the attainable in RNA genotype-phenotype mapping. *J. Theor. Biol.* 194, 491–515
97 Schuster, P. (2003) Molecular insights into evolution of phenotypes. In *Evolutionary Dynamics: Exploring the Interplay of Selection, Accident, Neutrality, and Function* (Crutchfield, J.P. and Schuster, P., eds), pp. 163–215, Oxford University Press