Design principles of multi-map variation in biological systems

Juan F Poyatos
Logic of Genomic Systems Laboratory, CNB-CSIC, Madrid 28049, Spain

Complexity in biology is often described using a multi-map architecture, where the genotype, representing the encoded information, is mapped to the functional level, known as the phenotype, which is then connected to a latent phenotype we refer to as fitness. This underlying architecture governs the processes that drive evolution. Moreover, natural selection, along with other neutral forces, can modify these maps. At each hierarchical level, variation is observed. Here, I propose the need to establish principles that can aid in understanding the transformation of variation within this multi-map architecture. Specifically, I will introduce three, related to the presence of modulators, constraints, and the channeling of variation. By comprehending these design principles in various biological systems, we can gain better insights into the mechanisms underlying these maps and their evolutionary dynamics.

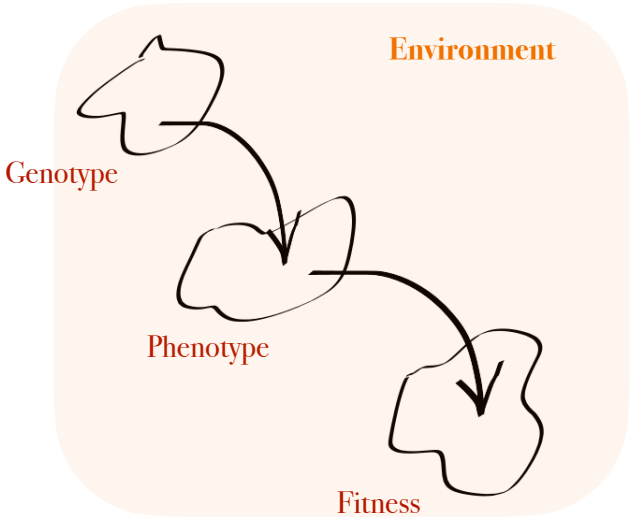

Figure 1: **Multi-map architecture of biological complexity**. The multi-map architecture is a fundamental framework for understanding biological complexity. It involves the mapping of genotype to phenotype and phenotype to fitness. The action of these mappings can be influenced by environmental factors, and in certain cases, the arrows can be reversed.

Introduction

It is well recognized that biology encompasses a multitude of mappings [1]. In a given environment, we observe a mapping from genotype to phenotype, representing function, and from phenotype to fitness, which can be perceived as a latent phenotype that is not readily observable but connected to selection (Figure 1). While it is acknowledged that some of these mappings may not be unidirectional, for the sake of simplicity, let's maintain that assumption.

Within this framework, we encounter variation at the genotypic, phenotypic, and fitness levels [2]. This variation leads us to three crucial characteristics frequently discussed in biology: robustness, plasticity, and evolvability [3]. These properties elucidate how variation at one level may not be experienced at all at a different one (robustness) or may be experienced as response to perturbations (plasticity) at different levels. They also highlight how variation drives evolution (evolvability) and how environmental factors modify all these aspects.

Below, I present an argument for the existence of design principles governing multi-map variation. While the ultimate generality of these principles may be uncertain, I advance that their presence can be observed across a wide range of biological systems. The search for such principles is an ongoing endeavor in other areas of biology and has contributed, among other things, to the development of a theory of biological networks [4, 5]. With this in mind, I propose three integral elements that contribute to what could eventually be understood as a general theory of multi-map variation. First, there exist "malleable" agents that control variation. Second, there are "rigid" constraints on variation. Third, we observe a "modular" channeling of variation.

Characterizing these principles in different biological systems will allow us to unravel the mechanisms underlying the mapping process, the forces that shape evolutionary trajectories, and the factors that influence the emergence of new traits. Ultimately, this knowledge can enhance our ability to interpret and predict evolutionary dynamics in diverse biological contexts. Let me explain these three elements further below.

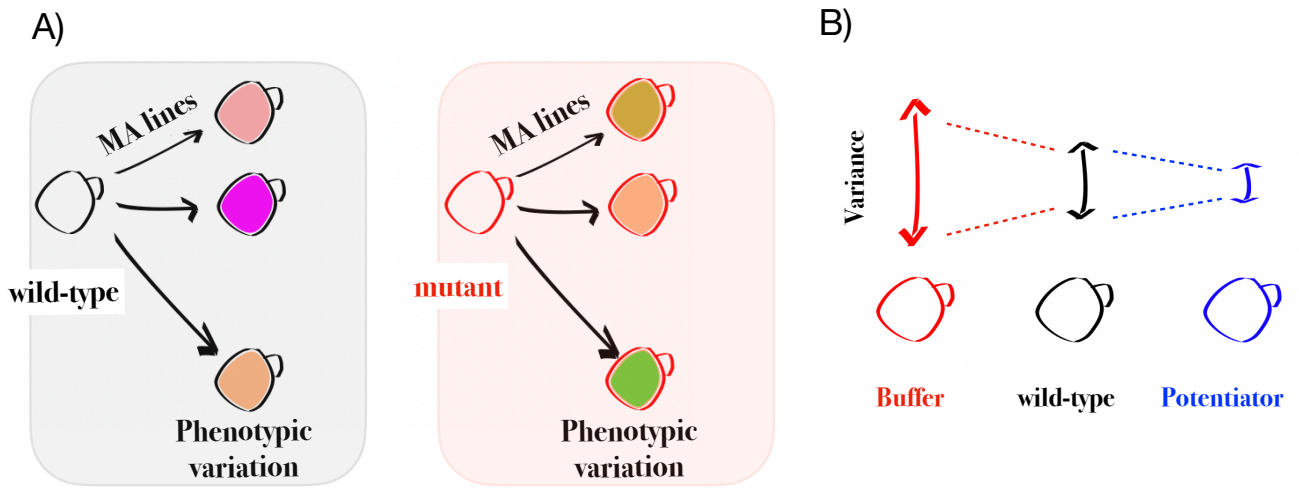

Figure 2: **Buffers and potentiators of variation control**. A) Each arrow represents a mutation accumulation (MA) line from a reference budding yeast (*Sacharomyces cerevisiae*) genotype (wild-type on the left and mutant on the right). Both wild-type and mutant experience the same MA trajectories. The filled color indicates the corresponding phenotype of that individual yeast. B) Mutants that exhibit an increase in variation, denoted by the vertical arrow, are referred to as buffers (shown in red). The absence of buffers results in an amplified variation. On the other hand, mutants that display a decrease in variation are referred to as potentiators (shown in blue).

Results

"Malleable" agents of variation control

Let's begin with the first component, the presence of malleable agents of variation control in biological systems. Imagine a scenario where a system undergoes multiple mutations. In more technical terms, we can consider a trajectory of mutation accumulation (MA), where mutations are allowed to accumulate over a given time period [6]. This represents the fundamental variation at the genotype level. My argument will be that there are elements within the system that modify the phenotypic variation generated as a consequence of these mutations.

To illustrate this concept, let's consider the budding yeast *Saccharomyces cerevisiae* as a model system. By following the MA trajectory, we can observe mutations occurring in the sequences encoding enzymes associated with its metabolism (Figure 2). Starting from the wild-type strain, we can generate multiple trajectories and quantify a fixed phenotype of the resulting mutants. Since we have numerous lines, we can compute the variation in this phenotype, such as its variance.

Now, let's modify the initial background by introducing a mutation in a single element of the system (from black to red in Fig. 2). From this new background, we generate the same MA lines and once again measure the phenotypic variation. In this case, the variance may either increase or decrease compared to the initial background, indicating that certain elements of the system, such as the enzymes in our example, function as modulators of variation.

It is important to note that this set of buffers and potentiators can change, and that the same agent may transition from acting as a buffer to becoming a potentiator or vice versa. Their classification as buffers or potentiators is contingent upon the specific operational regime of the system under consideration, emphasizing their malleability and contextual nature [7].

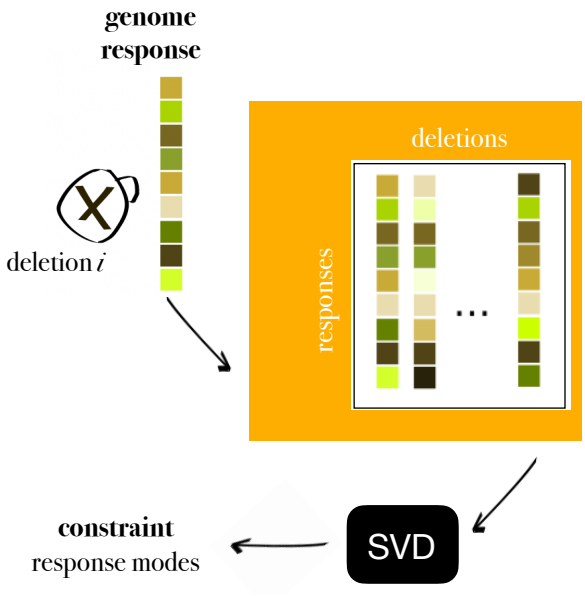

Figure 3: **Restrictions on variation**. The genome responses to single gene deletions were monitored in a collection of budding yeast mutants. This dataset forms a matrix of transcriptional responses (rows) versus deletions (columns). To reduce the dimensionality of this matrix, we employed singular value decomposition (SVD). We hypothesize that the resulting modes obtained through this technique represents constraints on the potential responses of the system, which may lead to suboptimality (see main text for details).

"Hard" restrictions on variation

The second element I present discusses the existence of constraints on variation. The term "constraint" has been a subject of historical discussions in biology, particularly regarding the generation of phenotypic variation, such as developmental constraints [8]. Nevertheless, my argument proceeds as follows. I consider specifically a dataset obtained from the budding yeast, which encompasses hundreds of gene knockouts (representing one-quarter of yeast genes) for which genome-wide mRNA expression was monitored [9]. This dataset can be represented as a matrix of gene expression *vs.* deletions (Figure 3).

To explore the underlying structure of this matrix, we utilize singular value decomposition (SVD), a technique similar to principal component analysis. SVD allows us to identify a set of patterns that we refer to as expression "modes". They represent recurring global expression changes, where groups of genes exhibit coordinated and consistent changes in their expression levels across multiple gene deletion strains. Thus, each transcriptional response can be deconstructed into a combination of these patterns, which I argue reflect the presence of constraints on variation.

An important insight derived from our research is that these modes do not necessarily drive "optimal" responses. Contrary to the expectations of optimality, we have discovered that the erroneous activation of certain genes, resulting from these constraints, can contribute to the fitness defects observed in deletions [10]. In other words, the presence of stringent restrictions on variation can generate suboptimal phenotypes. This highlights the importance of considering the interplay between constraints, variation, and fitness.

"Modular" channeling of variation

The third principle emphasizes how variation at the phenotypic level can be channeled in a modular manner to impact fitness. To investigate this, we focus on a class of mutations that we term "complex" mutations [11]. These mutations occur in molecular agents that exhibit high pleiotropy,

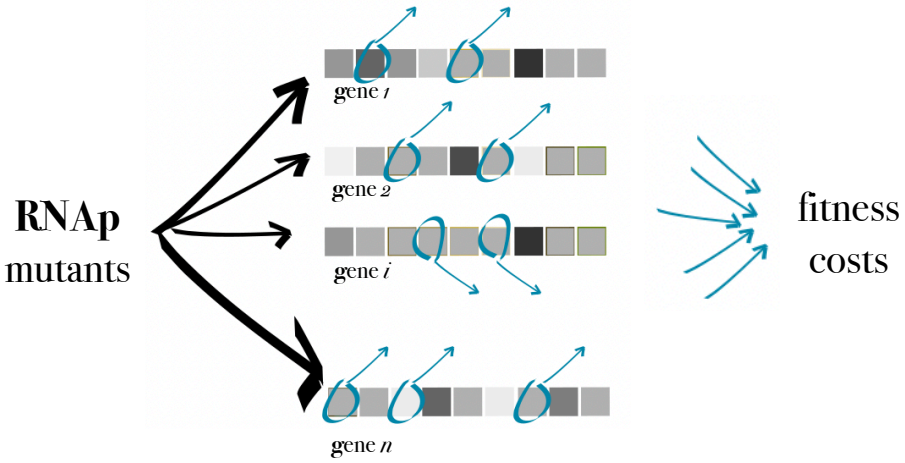

Figure 4: **Channeling on variation**. Mutations in a group of RNAp mutants result in alterations in various characteristics associated with the transcriptional response of constitutive genes (i.e., those not regulated by transcriptional factors). We quantified these features ("barcodes" in the figure) and observed that only a subset of the phenotypic variation contributes to fitness variation (blue circles/arrows). This indicates that phenotypic variation is channeled in a modular manner, ultimately shaping the variation observed at the fitness level.

such as the enzyme RNA polymerase (RNAp), with significant implications for phenotypes and their effect on fitness.

We investigated the impact of a specific set of mutants on gene expression, utilizing *Escherichia coli*, as experimental system. To do this, we quantified various phenotypic features closely associated with transcriptional efficiency and the alarmone ppGpp, which is a stress signaling system in bacteria (each box in Figure 4 represents one such feature; we assessed these features by monitoring promoter activity). It is essential to note that we used a set of constitutive genes to measure the expression response to the different RNAp mutations, excluding in this way the interference of additional regulatory signals, such as transcriptional factors. With this data in hand, we developed a model to understand how fitness changes in the mutants as a function of the observed phenotypic features. Interestingly, we found that only a subset of them significantly influences fitness, as measured by the growth rate [11].

In a broader context, this demonstrates how certain types of genetic variation give rise to extended phenotypic diversity. The phenotypic variation is represented by barcodes (Fig. 4), which indicate the different features associated with each gene. Notably, from this wide range of phenotypes, only a few contribute to fitness (blue circles in Fig. 4). This proposes a model that accounts for both extended phenotypic pleiotropy and fitness-relevant modularity [12]. In our context, the full combination of constitutive genes and their associated features demonstrates phenotypic pleiotropy. It becomes evident that only a subset of these phenotypic elements contributes to fitness, indicating fitness modularity. Consequently, fitness, as a complex trait, exhibits robustness to certain phenotypic changes.

Discussion

Three fundamental questions arise from the standard mapping architecture of biology (Figure 1). First, how did such architecture evolve? This question encompasses exploring the various developmental mechanisms contributing to phenotypes and understanding how phenotypes ultimately impact fitness. The second question delves into the generation of variation at each level of the architecture. This problem also refers to the balance between error generation and tolerance in

biology. Finally, the third problem, which I have examined here, investigates from a functional point of view the mapping of variation from one level to another.

Given the existence of multi-map variation, my focus is to identify generic features that underlie the transformation of variation across different levels. Modulators provide insights into the presence of elements that adjust the transition from genetic to phenotypic variation [7, 13]. These elements have the capacity to amplify or diminish the variation at the lower level, both aspects also related to ideas of network controlability (from complex systems [14]) and modifier genes (from genetics [15]).

While several molecular agents, such as the heat shock protein HSP90, have been identified as essential modulators [13], it is important to note that this characteristic is not inherently linked to the molecular attributes of the element itself. Furthermore, this modulatory role is not static and can vary over time [16]. Whether an element acts as a buffer or potentiator is contingent upon the distinct operational regime of the system being investigated [7].

The presence of constraints serves to highlight two important aspects. First, they indicate that biological responses, or behaviours, are both limited and can sometimes result in suboptimal situations [10]. This observation has been particularly emphasised in discussions surrounding multi-objective optimisation in biology, biological trade-offs, and related topics [17]. Secondly, the functioning of biological systems ultimately operates within a reduced number of dimensions, illustrating the pervasive nature of dimensional reduction [18, 19].

This becomes increasingly apparent as we have the ability to measure an expanding range of molecular phenotypes, such as through techniques like single-cell RNA sequencing. Despite the vast amount of variation present in these phenotypes, it is intriguing to observe that it can effectively be reduced to a few key dimensions. In some cases, these dimensions can be associated with dynamical attractors [20]. Recent discussions on how evolution contributes to this reduction in dimensionality are particularly relevant and merit consideration in this context [21]. Exploring the role of evolutionary processes in shaping the observed reduction in dimensionality can shed light on the mechanisms that drive the emergence of key phenotypic features and regulatory states.

Lastly, I emphasized the process of channeling variation to the fitness level. It is important to recognise that not all variation at the phenotypic level is necessarily relevant to fitness. This brings us back to the earlier comment I introduce regarding the balance associated with precision. It also evokes the concept of "sloppy" parameters in the context of biological models [22], where parameters with a wide range of values can yield similar outputs, specifically in relation to fitness.

This parallel draws attention to the notion that in biological systems, there is a certain tolerance for variation that does not significantly impact fitness. This flexibility allows for robustness and adaptability in the face of changing conditions or perturbations. It highlights the inherent complexity of the relationship between phenotypic variation and fitness outcomes, emphasising that not all variations are equally consequential for evolutionary success or fitness optimisation [23].

A different issue pertains to the origins of these principles. Following the spirit of Tinbergen's four questions (mechanism, development, function, and evolutionary history) [24], the aforementioned principles can be associated with the first three questions, but not the fourth. Exploring the selective or neutral processes that give rise to, for example, a molecule functioning as a malleable modulator of variation in a given condition is beyond the scope of this note. However, it presents an intriguing problem that warrants further investigation.

Lastly, I have illustrated these principles using three of my own works. This is not to imply that these works hold particular significance above others, as they do not (see, for instance, the very interesting special issue here [25]). Rather, they are examples with which I am intimately familiar, and through detailed study of these cases, I have been able to develop and extrapolate the proposed principles. My intention is that this commentary stimulates a more thoughtful discussion on how we can comprehend the "organised complexity" [26] of biology in the 21st century. Many things are at stake.

Acknowledgements

The majority of this manuscript was developed as an outcome of the workshop titled "Theory of Plasticity and Robustness in Natural and Artificial Systems" held at the Barcelona Collaboratorium, from July 3 to 5, 2023 and sponsored by the LifeHUB-CSIC network. I thank all the participant of this workshop for the stimulating discussions and Fernando Casares and Álvaro Sánchez for comments on an earlier version of the manuscript. This work was supported by grant PID2019-106116RB-I00 from the Spanish Ministry of Economy and Competitiveness with European Regional Development Fund. Comments are welcome.

References

[1] Richard C. Lewontin. *The genetic basis of evolutionary change*. Columbia biological series 25. New York: Columbia Univ. Pr, 1974.

[2] Benedikt Hallgrímsson and Brian Keith Hall, eds. *Variation*. Amsterdam ; Boston: Elsevier Academic Press, 2005.

[3] C.H. Waddington. *The Strategy of the Genes*. 0th ed. Routledge, Apr. 1957.

[4] Juan F. Poyatos. "On the Search for Design Principles in Biological Systems". In: *Evolutionary Systems Biology*. Ed. by Orkun S. Soyer. Vol. 751. New York, NY: Springer New York, 2012, pp. 183–193.

[5] Uri Alon. *An introduction to systems biology: design principles of biological circuits*. Second edition. Boca Raton London New York: CRC Press, Taylor & Francis Group, 2019.

[6] Daniel L. Halligan and Peter D. Keightley. "Spontaneous Mutation Accumulation Studies in Evolutionary Genetics". In: *Annual Review of Ecology, Evolution, and Systematics* 40.1 (Dec. 2009), pp. 151–172.

[7] Juan F. Poyatos. "Genetic buffering and potentiation in metabolism". In: *PLOS Computational Biology* 16.9 (Sept. 2020), e1008185.

[8] J. Maynard Smith et al. "Developmental Constraints and Evolution: A Perspective from the Mountain Lake Conference on Development and Evolution". In: *The Quarterly Review of Biology* 60.3 (Sept. 1985), pp. 265–287.

[9] Patrick Kemmeren et al. "Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors". In: *Cell* 157.3 (Apr. 2014), pp. 740–752.

[10] Károly Kovács et al. "Suboptimal Global Transcriptional Response Increases the Harmful Effects of Loss-of-Function Mutations". In: *Molecular Biology and Evolution* 38.3 (Mar. 2021), pp. 1137–1150.

[11] Pablo Yubero and Juan F Poyatos. "Dissecting the Fitness Costs of Complex Mutations". In: *Molecular Biology and Evolution* 38.10 (Sept. 2021), pp. 4520–4531.

[12] Grant Kinsler, Kerry Geiler-Samerotte, and Dmitri A Petrov. "Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation". In: *eLife* 9 (Dec. 2020), e61271.

[13] Rachel Schell, Martin Mullis, and Ian M. Ehrenreich. "Modifiers of the Genotype–Phenotype Map: Hsp90 and Beyond". In: *PLOS Biology* 14.11 (Nov. 2016), e2001015.

[14] J. Doyle. "Robust and optimal control". In: *Proceedings of 35th IEEE Conference on Decision and Control*. Vol. 2. Kobe, Japan: IEEE, 1996, pp. 1595–1598.

[15] Jesse D. Riordan and Joseph H. Nadeau. "From Peas to Disease: Modifier Genes, Network Resilience, and the Genetics of Health". In: *The American Journal of Human Genetics* 101.2 (Aug. 2017), pp. 177–191.

[16] Kerry A. Geiler-Samerotte et al. “Selection Transforms the Landscape of Genetic Variation Interacting with Hsp90”. In: *PLOS Biology* 14.10 (Oct. 2016), e2000465.

[17] O. Shoval et al. “Evolutionary Trade-Offs, Pareto Optimality, and the Geometry of Phenotype Space”. In: *Science* 336.6085 (June 2012), pp. 1157–1160.

[18] Mónica Chagoyen and Juan F. Poyatos. “Complex genetic and epigenetic regulation deviates gene expression from a unifying global transcriptional program”. In: *PLOS Computational Biology* 15.9 (Sept. 2019), e1007353.

[19] Kabir Husain and Arvind Murugan. “Physical Constraints on Epistasis”. In: *Molecular Biology and Evolution* 37.10 (Oct. 2020). Ed. by Claus Wilke, pp. 2865–2874.

[20] Sui Huang et al. “Cell Fates as High-Dimensional Attractor States of a Complex Gene Regulatory Network”. In: *Physical Review Letters* 94.12 (Apr. 2005), p. 128701.

[21] Jean-Pierre Eckmann and Tsvi Tlusty. “Dimensional reduction in complex living systems: Where, why, and how”. In: *BioEssays* 43.9 (2021), p. 2100062.

[22] Ryan N Gutenkunst et al. “Universally Sloppy Parameter Sensitivities in Systems Biology Models”. In: *PLoS Computational Biology* 3.10 (Oct. 2007), e189.

[23] Andreas Wagner. *Robustness and evolvability in living systems*. Princeton studies in complexity. Princeton, N.J. Oxford: Princeton University Press, 2007.

[24] N. Tinbergen. “On aims and methods of ethology”. In: *Tinbergen’s Legacy*. Ed. by Johan Bolhuis. 1st ed. Cambridge University Press, Jan. 2009, pp. 1–24.

[25] Vincent Debat and Arnaud Le Rouzic. “Canalization, a central concept in biology”. In: *Seminars in Cell & Developmental Biology* 88 (Apr. 2019), pp. 1–3.

[26] W Weaver. “Science and complexity”. In: *American Scientist* 36 (1948), pp. 536–544.