Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 39 (2021) 76–78

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Comment

The search for universality in evolutionary landscapes Comment on "From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics" by Susanna Manrubia, José A. Cuesta, et al.

Michael Manhart *, Sebastian Bonhoeffer

Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland

Received 19 August 2021; accepted 19 August 2021
Available online 23 August 2021
Communicated by E. Shakhnovich

The development of molecular biology in the mid-twentieth century revolutionized the life sciences by revealing how genetic information is stored in DNA sequences and expressed as RNA and protein sequences. Although the concept of populations evolving on fitness landscapes dates to Sewall Wright and the modern synthesis [1], the discovery that genotypes are encoded as long sequences of nucleotides created a new conundrum for evolutionary biology: Given the "hyper-astronomical" number of possible genotypes — more possibilities than atoms in the observable universe — how does evolution ever find functional states [2]? Unlike the empirical discoveries of molecular biology, this was a problem in need of theoretical solutions, since the enormous sizes of sequence spaces preclude exhaustive experimental or computational approaches. The resulting theory of evolution on fitness landscapes, or more generally genotype-phenotype (GP) maps, has largely solved this problem by demonstrating that these landscapes are neither flat nor random, but rather have characteristic structure that can both facilitate the evolution of new phenotypes as well as maintain existing phenotypes in the face of mutational deterioration.

In a new review, Manrubia, Cuesta, et al. [3] explore many of the key models and empirical data contributing to this theory. As they rightly emphasize, some of the most compelling models have been those that make *ab initio* predictions of phenotypes, such as RNA [4] or protein structures [5], from the underlying genotypes. Because these predictions are based on fundamental principles of physics and chemistry, these models plausibly capture generic features of landscapes across different organisms and environments. The last decade and a half has also brought an explosion of new experiments that directly measure properties of GP maps, allowing empirical verification of many model predictions. These experiments generally fall into two classes: combinatorially-complete measurements of a set of mutations in a small part of the genome [6,7]; and mutational scans that exhaustively sample single mutations (and sometimes double or triple mutations) on a reference genotype [8], forming a complete picture of the local neighborhood of a genotype.

These models and empirical data have revealed a number of important insights into the structure of GP maps and how evolutionary dynamics proceed on them. First and foremost is a solution to the initial problem posed by molec-

DOI of original article: https://doi.org/10.1016/j.plrev.2021.03.004.
* Corresponding author.
E-mail address: michael.manhart@env.ethz.ch (M. Manhart).

https://doi.org/10.1016/j.plrev.2021.08.004

ular biology: new phenotypes are often highly evolvable on GP maps, with single mutations leading to incremental improvements that can be positively selected [2]. At the same time, these landscapes exhibit significant robustness, meaning that phenotypes can be stable to mutational perturbations over evolution. A major accomplishment of models has been to elucidate the relationship between evolvability and robustness. In particular, we have learned that high evolvability can actually go hand-in-hand with high robustness for phenotypes, despite their apparent contradiction [9]. Underlying this result is a concept known as phenotypic bias, meaning that some phenotypes are produced by many more genotypes than are other phenotypes; for example, the most common secondary structure of a 20 nucleotide-long RNA is represented by 10 orders of magnitude more genotypes than is the least common structure [10]. This means that the evolved phenotypes we observe in extant organisms are driven not purely by selection for those phenotypes, but also by an entropic tendency that favors phenotypes that are common across genotype space. This conceptual result can explain a number of important observations in molecular evolution, such as the marginal stability of proteins [11], the prevalence of loss-of-function mutations in evolution experiments [12], and the existence of molecular cross-talk between proteins [13] and between proteins and DNA regulatory sites [14].

One of the most exciting possibilities raised by this work has been the existence of universal, effective models that capture the essential features of a wide range of landscapes without relying on details of specific systems. For example, a genotypic sequence can be partitioned into a subset of constrained loci, which contribute significantly to phenotypes under selection, and a subset of unconstrained or neutral loci [15]. Models constructed on this feature alone are sufficient to make statistical predictions, such as the distribution of neutral networks of genotypes, that quantify phenotypic bias [16]. There are also various models capturing the essential statistical features of other properties of GP maps such as epistasis [17,18].

How far can we take this approach toward the next frontiers of biology? While models and experiments for GP maps in the context of well-defined molecular phenotypes — such as RNA secondary structure or a protein binding a ligand — have driven much of this success, Manrubia, Cuesta, et al. point out that there is still a large gap between these molecular phenotypes and properties of whole cells or organisms, especially the organismic fitness that determines whether a lineage expands or contracts in a population. Multiscale models have been constructed to address this problem, but these typically require strong assumptions about the mechanisms by which molecular phenotypes give rise to cell growth and division. Experiments, on the other hand, do often measure organismic fitness, but the inevitable limitation of these experiments by the underlying combinatorics of genotype space means that we will eventually need models and theoretical approaches to fully address this problem.

One appealing approach to the problem of phenotype-fitness maps is to focus on effective statistical models, as aforementioned for genotype-phenotype maps, to uncover universal properties. For example, one can measure (or infer) a large number of phenotypes for a collection of genotypes and determine the quantitative map between these phenotypes and fitness. So far this approach has been used mainly to determine the number of independent phenotypes contributing to fitness [19,20] and the modularity of the phenotype-phenotype interaction network [21]; for example, one striking result from these studies has been that the dimensionality of fitness-relevant phenotypes is $\sim$10. However, a major challenge going forward is to determine whether the properties of phenotype-fitness maps can ever be generalized across environmental conditions: while molecular phenotypes often have more straightforward dependence on the environment (e.g., through temperature or ligand concentration), complex phenotypes, especially fitness, may depend on the environment idiosyncratically. This problem is especially acute for microbes, which exist in dynamic ecosystems consisting of many different species constantly altering their own chemical and metabolic environment. For example, mutations that change the metabolic secretions of microbial cells can “deform” the fitness landscape experienced by future mutations, leading to non-commutativity between these mutations [22]. The promise of landscapes for predicting evolutionary biology will hinge on our ability to glean universal principles that account for these complexities in natural biological systems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

MM was supported by an Ambizione Fellowship from the Swiss National Science Foundation (PZ00P3_180147).

References

[1] Wright S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proc 6th Int Congr Genet, vol. 1. 1932. p. 356–66.
[2] Maynard Smith J. Natural selection and the concept of a protein space. Nature 1970;225:563–4.
[3] Manrubia S, Cuesta JA, Aguirre J, Ahnert SE, Altenberg L, Cano AV, et al. From genotypes to organisms: state-of-the-art and perspectives of a cornerstone in evolutionary dynamics. Phys Life Rev 2021;38:55–106.
[4] Schuster P, Fontana W, Stadler PF, Hofacker IL. From sequences to shapes and back: a case study in RNA secondary structures. Proc R Soc B 1994;255:279–84.
[5] Li H, Helling R, Tang C, Wingreen N. Emergence of preferred structures in a simple model of protein folding. Science 1996;273:666–9.
[6] Weinreich DM, Delaney NF, DePristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 2006;312:111–4.
[7] Podgornaia AI, Laub MT. Pervasive degeneracy and epistasis in a protein-protein interface. Science 2015;347:673–7.
[8] Firnberg E, Labonte JW, Gray JJ, Ostermeier M. A comprehensive, high-resolution map of a gene's fitness landscape. Mol Biol Evol 2014;31:1581–92.
[9] Wagner A. Robustness and evolvability: a paradox resolved. Proc R Soc B 2008;275:91–100.
[10] Schaper S, Louis AA. The arrival of the frequent: how bias in genotype-phenotype maps can steer populations to local optima. PLoS ONE 2014;9:e86635.
[11] Taverna DM, Goldstein RA. Why are proteins marginally stable? Proteins 2002;46:105–9.
[12] Behe MJ. Experimental evolution, loss-of-function mutations, and "the first rule of adaptive evolution". Q Rev Biol 2010;85:419–45.
[13] Deeds EJ, Ashenberg O, Shakhnovich EI. A simple physical model for scaling in protein-protein interaction networks. Proc Natl Acad Sci USA 2006;103:311–6.
[14] Friedlander T, Prizak R, Guet CC, Barton NH, Tkačik G. Intrinsic limits to gene regulation by global crosstalk. Nat Commun 2016;7:12307.
[15] Greenbury S, Ahnert S. The organization of biological sequences into constrained and unconstrained parts determines fundamental properties of genotype-phenotype maps. J R Soc Interface 2015;12:20150724.
[16] Manrubia S, Cuesta JA. Distribution of genotype network sizes in sequence-to-structure genotype-phenotype maps. J R Soc Interface 2017;14:20160976.
[17] Szendro IG, Schenk MF, Franke J, Krug J, de Visser JAGM. Quantitative analyses of empirical fitness landscapes. J Stat Mech 2013;2013:P01005.
[18] Reddy G, Desai MM. Global epistasis emerges from a generic model of a complex trait. eLife 2021;10:e64740.
[19] Wytock TP, Motter AE. Predicting growth rate from gene expression. Proc Natl Acad Sci USA 2019;116:367–72.
[20] Maeda T, Iwasawa J, Kotani H, Sakata N, Kawada M, Horinouchi T, et al. High-throughput laboratory evolution reveals evolutionary constraints in Escherichia coli. Nat Commun 2020;11:5970.
[21] Kinsler G, Geiler-Samerotte K, Petrov DA. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. eLife 2020;9:e61271.
[22] Bajić D, Vila JCC, Blount ZD, Sánchez A. On the deformability of an empirical fitness landscape by microbial evolution. Proc Natl Acad Sci USA 2018;115:11286–91.