A Revival of the Landscape Paradigm

Large Scale Data Harvesting Provides Access to Fitness Landscapes

T he first large-scale fitness analysis of HIV in presence and absence of antiviral drugs has been presented in a recent publication [1]. This study confirms expectations, in particular ruggedness and neutrality of the HIV fitness landscape, but it provides at the same time also unexpected results like long-range correlations and surprisingly little influence of antiviral agents on the landscape structure. Here, we make use of this occasion to present a brief account on the development of the landscape concept from a pure metaphor to an experimentally assisted tool for modeling and understanding evolution.

Sewall Wright [2], one of the three great scholars of population genetics conceived the idea of a fitness landscape as a metaphor for the evolutionary process: species or subspecies are occupying local fitness optima in a rugged landscape (Fig. 1). Fitness is plotted upon genotype space, which is a high-dimensional space with individual genotypes differing in the recombination pattern of genes as elements. The metaphor was introduced as a visualization of his theory of shifting balance as a model of evolution that in a nutshell may be formulated as a process in three logical steps [3, 4]: (i) Random drift leads to semi-isolated subpopulations or demes within the global population, which are losing fitness because of accidental loss of fittest genotypes known as Muller's ratchet [5], (ii) mass selection[1] acts on complex genetic interaction networks and raises the fitness of subpopulations, and (iii) interdemic selection raises the fitness of the global population. Environmental change shifts the adaptive peaks on the landscape and drives evolutionary dynamics. Clearly, the landscape concept is suggestive for the Wright model and facilitates its understanding.

Wright's model of evolution has been heavily criticized by Ronald Fisher and others. Apart from the practical impossibility to measure fitness in the first half of the 20th century and the apparent lack of knowledge on the nature of genes and genotypes that rendered void any attempt to quantify fitness on landscapes, Fisher's and Wright's views on evolution differed in many aspects. Fisher's concept called large population size theory assumes that (i) the ecological context of evolution is large panmictic populations[2] rather than small subpopulations, (ii) the

PETER SCHUSTER

Peter Schuster, Editor-in-Chief of Complexity is at the Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, 1090 Wien, Austria (e-mail: pks@tbi.univie.ac.at)

---

[1]*Mass selection is selection of phenotypes and can be defined as breeding stock of those members of a population exhibiting desirable qualities or elimination of those showing undesirable qualities in artificial selection.*

[2]*A panmictic population is unstructured and reproduction occurs by random mating.*
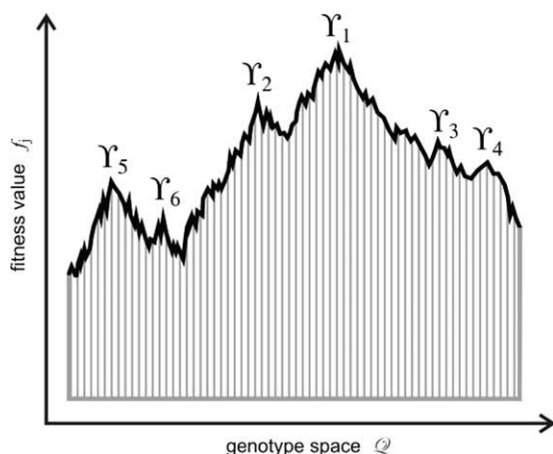
A sketch of a fitness landscape in the sense of Sewall Wright's metaphor. Populations Υ occupy local optima in a rugged fitness landscape. The support—here only a domain on the abscissa axis—is multidimensional. In the original article by Wright [2], the fitness landscape was built upon the genotype recombination space, which is huge with respect to the number of possible genotypes and has a large number of dimensions. If mutation is considered as the source of variation of genotypes the support of the landscape is sequence space that is again huge and high-dimensional.

major process driving evolutionary change is mutation and natural selection rather than a combination of local natural selection, random genetic drift, migration and interdemic selection, (iii) the genetic basis of change is additive genetic effects and context independence of allelic functions rather than epistasis and pleiotropy leading to context dependent alleles, and (iv) the ultimate driving force for evolution is refinement of existing adaptations in a stable or slowly changing environment rather than adaptive novelty in constantly changing environments. Fisher's basic concept does not require the landscape metaphor nor does it gain in transparency from it, and consequently Fisher rejected and belittled Wright's picture.

Wright borrowed his landscape idea from physics where the concept of a potential energy surface is used to determine, for example, the motion of particles. In quantum mechanics the potential energy surface got its precise meaning by the work of Max Born and Robert Oppenheimer in the theory of molecules [6]. Their approach, known as Born–Oppenheimer approximation, became the basis for quantum mechanics of molecules: Fast motion of electrons is separated from slow motion of the atomic nuclei, and the potential energy surface is the connection between both. In other words, the nuclei move so slowly so that they do not feel the potential of the individual moving electronic charges but a time averaged smeared potential that encapsulates also the quantum effects of electron motion:

$$(T_e + V(\mathbf{R}))\Psi_k = E_k(\mathbf{R})\Psi_k \text{ and}$$
$$(T_k + E_k(\mathbf{R}))\Xi_l = E_{kl}\Xi_l \quad (1)$$

The electronic Schrödinger equation (left-hand side) and the Schrödinger equation for nuclear motion (right-hand side) are coupled via the potential energy surface $E_k(\mathbf{R})$ referring to the kth electronic state of the molecule, T_e, T_k, and $V(\mathbf{R})$ are the Hamilton operators for the kinetic energies of electrons and nuclei and for the potential energy, respectively, Ψ_k and Ξ_l the wave functions for electrons and nuclei, and E_{kl} eventually the energy eigenvalue of the kth electronic state and the lth state of nuclear motion. Since the positions of the atomic nuclei enter the electronic Hamilton operator, $T_e + V(\mathbf{R})$, via the potential energy, $V(\mathbf{R})$ constitutes the second coupling term, which is rather trivial because the atomic nuclei are practically at rest for the moving electrons. The potential energy surface $E_k(\mathbf{R})$ of a general molecule has dimension $3N - 6$ where N is the number of atoms in the molecule and accordingly, the dimension can be very large for macromolecules. Although firmly rooted in quantum physics, the calculation of multidimensional potential energy surfaces was practically impossible until the spectacular development of computer facilities and more efficient algorithms changed the situation in the nineteen hundred eighties: Energy surfaces of medium-size molecules (with about 10–100 atoms) are now accessible through numerical calculations and they are used routinely in molecular spectroscopy and chemical reaction dynamics.

Ab initio calculations of energy surfaces are not possible at present for larger systems from medium-sized molecules up to biological macromolecules but the potential energy surface concept is applied in the form of a very useful empirical heuristic often addressed as molecular mechanics or molecular dynamics [7–9]: The potential energy surface $E_k(\mathbf{R})$ is modeled by means of suitable functions with empirical parameters and molecular properties are derived by application of Newtonian mechanics, either by determination of molecular conformations through searches for minima of the

potential energy surface or by simulation of molecular motions through solving Newton's equations. The landscape concept turned out to be particularly useful for understanding of protein folding [10, 11]. Conformational landscapes are in widespread use and their predictive power becomes better and better mainly for two reasons: (i) the computational techniques are improved and (ii) the fast expanding collection of high-resolution structures in data bases provides a wealth of quantitative empirical knowledge. Nowadays, protein potential energy functions are also used for computational design of proteins [12].

The fast acceptance of the landscape idea and the remarkable progress in applications of various kinds, which took place in chemistry and in structural biology, are somewhat contrasting the development in evolutionary biology. Some theoretical models for fitness landscape construction assign values at random [13, 14], and others make use of known structures and functions of biomolecules, in particular RNA [15, 16]. The RNA model—derived from biopolymer structures and functions—has shown two features of landscapes built upon sequence space: (i) ruggedness and (ii) neutrality. The correlation length of structures turned out to be rather short [17]. It is worth illustrating short and long correlation lengths by means of mountain regions: the dolomites have small correlation lengths but Mount Fuji and other typical volcanoes like the inactive ones in the Massif Central in France show correlations over long distances. Until the advent of modern techniques in molecular genetics, almost nothing was known on global properties of experimental fitness landscapes, and apart from specific fitness information on local environments in sequence space also the knowledge on the consequences of mutations on fitness was *tabula rasa*. Fitness of organisms is a highly complex function and has many inputs, which are difficult to

control. Reducing the complexity of evolving systems, however, has led to simple replication assays that can be studied by the conventional tools of physical chemistry [18, 19]. RNA molecules with specific recognition sites are replicated in a solution with activated nucleotides and an enzyme for replication—a virus specific RNA-dependent RNA polymerase present—in excess. A well-developed theory of evolution built upon chemical reaction kinetics has been developed [20, 21] and the chemical reaction kinetics of extracellular RNA replication were studied in great detail [22–24]. This process is now as well understood as other chemical multistep reactions. Mutation being inevitable because of the low accuracy of the replicating enzyme [25] provides the source of innovation for cell-free evolution. Mutation may consist in single point mutations or in larger changes of the genetic information. Fitness in the case of in vitro evolution is a well-understood function of several parameters among them the RNA replication rate parameter and the binding parameter of the enzyme to the RNA [26] or in other words, fitness is readily accessible through measurements of the appropriate physical quantities.

Fitness and mutation play a dominant role in virus research and in the development of antiviral strategies [27–29]. (For a recent account on lethal mutagenesis of viruses see, for example, a recent "simply complex" [30]). The new techniques of molecular genetics, in particular cheap and fast sequencing and on chip technology, allow for data harvesting on large scales, for example, large numbers of random mutations in single-stranded DNA and RNA bacteriophages were collected in a study of fitness effects [31]. Nevertheless, particular care is needed when relatively small fitness differences are to be interpreted [32]. New data are coming up nowadays with a breath-taking pace. One of the latest large scale investigations on

HIV-1 has been mentioned initially. The study is based on predictive models for fitness measured through in vitro replication of HIV and in essence, it confirms the results of biopolymer models consisting in ruggedness and considerable neutrality. The correlation lengths of fitness values in this study is longer than expected and, what makes the investigation particularly interesting, the landscape for independent gene actions and the landscape for epistatic effects[3] are substantially more rugged and have shorter correlation lengths than the results obtained by combining both. All systems observed in nature are the result of long-time evolution and adaptation by selection. Accordingly, the interplay between the effects of independent gene action and epistasis might easily have been exploited to yield an optimal landscape for optimization.

Finally, we discuss a class of problems where landscapes have to be applied with care provided usage can be recommended at all. Dynamical fitness landscapes are frequently applied as metaphor for an illustration of co-evolution as well as for a visualization of the coupling between adaptation and environmental change [33]. The fitness landscape varies with time and populations adapt to this change by mutation and/or recombination and selection. For a comparison of dynamical fitness landscapes with the concept of a dynamic potential in molecular physics as shown in Eq. (1) it is important to distinguish carefully the the two different potentials in the equation: (i) The potential in the Hamilton operator for electronic motion $V(\mathbf{R})$, which contains the coordinates of the nuclei subsumed in $\mathbf{R}$, and (ii) the energy surface $E_k(\mathbf{R})$ that constitutes the conformational energy landscapes for nuclear motion that has been

---

[3]*Epistasis is the name for the effects of gene interactions relative to the independent gene model.*

already discussed. The work of Born and Oppenheimer on molecular motion provides a perfectly understood case of successful decoupling of processes on two different time scales. Nuclei are more than three orders of magnitude heavier than electrons and hence nuclear motion occurs on a time scale that is about the same factor slower. The electrons move very fast and hence electronic motion is quasi at equilibrium for the moving nuclei or vice versa, the electrons "see" the atomic nuclei practically at rest in their contribution to $V(\mathbf{R})$. Successful decoupling of processes is only possible when they occur on sufficiently different time scales. Environmental change, climate change for example, may be very slow and then species can adapt to a gradually varying landscape as it has happened regularly in the history of life on Earth and still is happening—

the metaphor is useful and provides insight into the evolutionary mechanism. In co-evolution of species, however, the two processes of mutual adaptation will commonly occur at the same time scale, since both species evolve through selection operating on the results of recombination and mutation. My claim is that the landscape metaphor becomes obsolete in such situations and I illustrate by means of an example: A hiker is climbing upward. How could he possibly reach the top if the landscape changes at the same speed? For example, if he wants to take a step upward but at the same instant, the landscape has changed and causes him to go downward, he will never reach the top. Virus adaptation is much faster than the evolution of the host organisms (consider, as an example, the co-evolution of a bacterium with its parasite [34]) but at least

in all higher organisms the virus population has to cope with the immune system rather than with the evolutionary change of the species and the immune system is able to respond fast. Again, the two time scales are very similar. By the same token the landscape metaphor is doomed to fail in case of fast environmental change. A proper description then is modeling symbiosis or arms races directly as dynamical systems and leaving the landscape concept aside.

In summary, unless applied to co-evolutionary situations that are not compatible with the use of landscapes, Wright's metaphor has become a useful tool for quantitative understanding of evolution and the modern data harvesting techniques provide a rich empirical resource, which can be used for the successful construction of specific fitness landscapes.

REFERENCES

1. Kouyos, R.D.; Leventhal, G.E.; Hinkley, T.; Haddad, M.; Whitcomb, J.M.; Petropoulos, C.J.; Bonhoeffer, S. Exploring the complexity of the HIV-1 fitness landscape. PLoS Genetics 2012, 8, e1002551.
2. Wright, S.The role of mutation, inbreeding, crossbreeding, and selection in evolution. In: International Proceedings of the Sixth International Congress on Genetics, Vol. 1; Jones, D.F., Ed.; Brooklyn Botanical Garden: Ithaca, New York, 1932; pp 356–366.
3. Skipper, R.A. Jr. The persistence of the R.A. Fisher – Sewall Wright controversy. Biology and Philosophy 2002, 17, 341–367.
4. Wade, M.J.; Goodnight, C.J. The theories of Fisher and Wright in the context of metapopulations: When nature does many small experiments. Evolution 1998, 52, 1537–1553.
5. Muller, H.J. The relation of recombination to mutational advance. Mutation Research 1964, 1, 2–9.
6. Born, M.; Oppenheimer, R. Zur Quantentheorie der Moleküle. Annalen der Physik 1927, 84, 457–484.
7. Allinger, N.L.; Burkert, U. Molecular Mechanics; American Chemical Society: Washington, DC, 1982.
8. Mackerell, A.D. Jr. Empirical force fields for biological macromolecules: Overview and issues. J Comput Chem 2004, 25, 1584–1604.
9. Das, R.; Baker, D. Macromolecular modeling with Rosetta. Annu Rev Biochem 2008, 77, 363–382.
10. Onuchic, J.N.; Luthey-Schulten, Z.; Wolynes, P.G. Theory of protein folding: The energy landscape perspective. Annu Rev Phys Chem 1997, 48, 545–600.
11. Wolynes, P.G. Energy landscapes and solved protein-folding problems. Phil Trans Roy Soc A 2005, 363, 453–467.
12. Suárez, M.; Jaramillo, A. Challenges in the computational design of proteins. J Roy Soc Interface 2009, 6, S477–S491.
13. Kauffman, S.; Levin, S. Towards a general theory of adaptive walks on rugged landscapes. J Theor Biol 1987, 128, 11–45.
14. Kauffman, S.A.; Weinberger, E.D. The N-k model of rigged fitness landscapes and its application to the maturation of the immune response. J Theor Biol 1989, 141, 211–245.
15. Schuster, P.; Fontana, W.; Stadler, P.F.; Hofacker, I.L. From sequences to shapes and back. A case study in RNA secondary structures. Proc Roy Soc B 1994, 255, 279–284.
16. Fontana, W.; Schuster, P. Continuity in evolution. On the nature of transitions. Science 1998, 280, 1451–1455.
17. Fontana, W.; Konings, D.A.M.; Stadler, P.F.; Schuster, P. Statistics of RNA secondary structures. Biopolymers 1994, 33, 1389–1404.
18. Spiegelman. S. An approach to the experimental analysis of precellular evolution. Q Rev Biophys 1971, 4, 213–253.
19. Biebricher, C.K. Darwinian selection of self-replicating RNA molecules. Evol Biol 1983, 16, 1–52.
20. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften 1971, 58, 465–523.
21. Eigen, M.; McCaskill, J.; Schuster, P. The molecular quasispecies. Adv Chem Phys 1989, 75, 149–263.
22. Biebricher, C.K.; Eigen M.; Gardiner, W.C. Jr. Kinetics of RNA replication. Biochemistry 1983, 22, 2544–2559.
23. Biebricher, C.K.; Eigen M.; Gardiner, W.C. Jr. Kinetics of RNA replication: Plus-minus asymmetry and double-strand formation. Biochemistry 1984, 23, 3186–3194.

24. Biebricher, C.K.; Eigen M.; Gardiner, W.C., Jr. Kinetics of RNA replication: Competition and selection among self-replicating RNA species. Biochemistry 1985, 24, 6550–6560.
25. Castro, C.; Arnold, J.J.; Cameron, C.E. Incorporation fidelity of the viral RNA-dependent RNA polymerase: A kinetic, thermodynamic, and structural perspective. Virus Res 2005, 107, 141–149.
26. Biebricher, C.K. Quantitative analysis of mutation and selection in self-replcating RNA. Adv Space Res 1992, 12, 191–197.
27. Domingo, E.; Holland, J.J. RNA virus mutations and fitness for survival. Annu Rev Microbiol 1997, 51, 151–178.
28. Domingo, E., Ed. Virus entry into error catastrophe as a new antiviral strategy. Virus Res 2005, 107, 115–228.
29. Bull, J.J.; Sanjuan, R.; Wilke, C.O. Theory of lethal mutagenesis for viruses. J Virol 2007, 81, 2930–2939.
30. Schuster, P. Lethal mutagenesis, error thresholds, and the fight against viruses. Rigorous modeling is facilitated by a firm physical background. Complexity 2011, 17(2), 5–9.
31. Domingo-Calap, P.; Cuevas, J.M.; Sanjuán, R. The fitness effect of random mutations in single-stranded DNA and RNA bacteriophages. PLoS Genetics 2009, 5, e1000742.
32. Bull, J.J.; Heineman.; Wilke, C.O. The phenotype-fitness map in experimental evolution of phages. PLoS One 2011, 6, e27796.
33. Kauffman, S.A. 1993. The Origins Order. Self-Organization and Selection in Evolution. Oxford University Press: New York, p243ff.
34. Kashiwagi, A.; Tetsuya, Y. Ongoing phenotypic and genotypic changes in experimental coevolution of RNA bacteriophage Qβ and Escherichia coli. PLoS Genetics 2011, 7, e1002188.