Progress in Biophysics and Molecular Biology 111 (2013) 75–82

Contents lists available at SciVerse ScienceDirect

Progress in Biophysics and Molecular Biology

journal homepage: www.elsevier.com/locate/pbiomolbio

Original research

From sequence to consequence and back

Stig W. Omholt*

Centre for Ecological and Evolutionary Synthesis, Department of Biology, University of Oslo, P.O. Box 1066, Blindern, N-0316 Oslo, Norway

ARTICLE INFO

Article history:
Available online 25 September 2012

Keywords:
Genotype–phenotype relation
Open physical systems
Systems dynamics
Self-transcendence
Morphospace
Genetic concepts
Biological order
Systemization
Burden

ABSTRACT

The genotype–phenotype relation is at the core of theoretical biology. It is argued why a mathematically based explanatory structure of this relation is in principle possible, and why it has to embrace both sequence to consequence and consequence to sequence phenomena. It is suggested that the primary role of DNA in the chain of causality is that its presence allows a living system to induce perturbations of its own dynamics as a function of its own system state or phenome, i.e. it capacitates living systems to self-transcend beyond those morphogenetic limits that exist for non-living open physical systems in general. Dynamic models bridging genotypes with phenotypic variation in a causally cohesive way are shown to provide explanations of genetic phenomena that go well beyond the explanatory domains of statistically oriented genetics theory construction. A theory originally proposed by Rupert Riedl, which implies that the morphospace that is reachable by the standing genetic variation in a population is quite restricted due to systemic constraints, is shown to provide a foundation for a mathematical conceptualization of numerous evolutionary phenomena associated with the phenotypic consequence to sequence relation. The paper may be considered a call to arms to mathematicians and the mathematically inclined to rise to the challenge of developing new formalisms capable of dealing with the deep defining characteristics of living systems.

1. Introduction

In his thought-provoking book *Life Itself* (Rosen, 1991), Robert Rosen asks: "What is it that enables living things, apparently so moist, fragile and evanescent, to persist while towering mountains dissolve into dust, and the very continents and oceans dance into oblivion and back?" We are still far from being able to give a full answer to this question. However, I am quite confident that it will encompass a deep understanding of the genotype–phenotype[1] relation phrased in mathematical terms. In the following I will outline why I think such a mathematically based explanatory structure is in principle possible, and why it has to embrace both sequence to phenotypic consequence and phenotypic consequence to sequence phenomena.

William Bateson coined the term *genetics* in a letter to Adam Sedgwick in 1905. In 1906 he announced the term publicly in his inaugural address *The Progress of Genetic Research* to the *Third International Conference of Hybridization and Plant Breeding* (Punett, 1928): "I suggest for the consideration of this conference the term Genetics, which sufficiently indicates that our labours are devoted to the elucidation of the phenomena of heredity and variation: in other words, to the physiology of descent, with implied bearing on the theoretical problems of the evolutionist and the systematist, and application to the practical problems of breeders whether of animals or plants."

The above quotation is a rare example of a deliberate formulation of the goals of a scientific discipline under establishment, and *genetics* is still defined as the science of heredity and variation in living organisms. As *heredity* denotes the transfer of characteristics from parent to offspring through their genes, and *variation* denotes the change in the form, position, state, or quality of something, it follows that one of the major goals of genetics is to account for observed biological patterns in the wide sense, within and across species, past and present, which in one way or another can be related to hereditary units and principles. Through the phrase "physiology of descent" Bateson apparently envisioned that it fell to

* Present address: Faculty of Medicine, Norwegian University of Science and Technology, P.O Box 8905, N-7491 Trondheim, Norway. Tel.: +47 90 94 09 85.
E-mail address: stig.omholt@ntnu.no.

[1] The terms *genotype* and *phenotype* were introduced by the Danish plant physiologist and geneticist Wilhelm Johannsen in 1909. An individual's *genotype* denotes the constitution of parts or all of its genetic material, while its *phenotype* may comprise anything from a single observable characteristic or trait to all conceivable ones. Thus any morphological, developmental, biochemical or physiological property all the way down to the subcellular level (including epigenetic marks), as well as any of the individual's behaviour and products of behaviour, is a phenotypic characteristic and belongs to the individual's *phenome*.

0079-6107/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.pbiomolbio.2012.09.003

genetics to explain how hereditary units actually cause observed phenotypic patterns in terms of physiological mechanisms. This is supported by the fact that in his book "Mendel's Principles of Heredity" (Bateson, 1909), he makes several attempts to actually establish an explanatory bridge between hereditary units and phenotypes by referring to physiological and physical principles.

When it comes to theoretical population and quantitative genetics, one hundred years of work is still captured by Richard Lewontin's description in his *Genetic Basis of Evolutionary Change* almost 40 years ago (Lewontin, 1974). Here he argued how population genetics has been mainly focused on genotype space and the development of mathematical machinery capable of describing changes in this space based on mutation, selection and genetic drift while quantitative genetics has been mainly focused on phenotype space and the development of statistical machinery capable of describing changes in this space based on variance component analysis. By operating in just one space, population genetics has had to make a caricature of phenotype space, while quantitative genetics has had to make a caricature of genotype space. Needless to say, the reason for this situation is the lack of understanding of the genotype–phenotype relation.

Better understanding of the genotype–phenotype relation is the key to a mature genetics theory *sensu* Bateson, capable of linking genotypes, phenotypes and population level genetic phenomena through a causal understanding of biological mechanisms. In order to address the whole range of evolutionary phenomena we need to acknowledge that causality associated with this relation flow in both directions. A theoretical biology, i.e. a theory of organisms, to be taken seriously, must be built upon the genotype–phenotype relation. I firmly believe that the major tool for raising this building is biophysically based systems biology in the wide sense combined with technology enabling massive experimental manipulation and measurement. This work has barely started, but it is the road we have to follow if we are to succeed in meeting Rosen's challenge.

2. The role of DNA in the chain of causality

As the genotype–phenotype relation is of such fundamental importance, and assuming that there is no substitute for mathematics in describing its essence, it is worthwhile to briefly contemplate how we should perceive the function of DNA in a mathematical explanatory framework of biological form in the wide sense.

Considering the development of the gene concept over the last hundred years, from an abstract entity though something coding for proteins and then to a curiously elusive object (Beurton and Falk, 2000), it seems appropriate to declare the gene concept to be dysfunctional as a scientific concept because of its ontological fuzziness. The genotype concept is in quite another situation because it has a direct physical interpretation in terms of DNA and there is no particular additional mechanism or effect attributed to it. A genotype may constitute everything from a single base pair to the whole genome of an individual, thus the concept provides the flexibility needed to span the whole spectrum of relationships between DNA information and phenotypic variation.

As several people have advocated over many years, DNA is among the most inert and nonreactive of organic molecules. It does not self-replicate and it does not make or do anything in any meaningful sense. And thus it cannot be considered to be a cellular sub-system. This is precisely why it is so biologically useful, and why the cellular machinery works so hard to prevent it from disintegrating (Shapiro, 2011). The terms *database* and *information organelle* are frequently used as metaphors for describing the function of genomic DNA and the cell nucleus, respectively. Since metaphors are very important instruments for thinking, their appropriateness must be evaluated in terms of how well they can be used to reach new understanding through the inferential capacity of the meanings and associations we attach to them. A database may be defined as a comprehensive collection of related data (i.e. pieces of information) organized for convenient access. If we understand information in the restricted technical sense as a sequence of symbols that can be interpreted as a message, the database and the information organelle metaphors are biologically sound. But in my view it is not very useful to just denote DNA as an information storage medium when it comes to assessing its role in the chain of causality in systems theoretical terms.

If we interpret information somewhat loosely as any kind of event that affects the state of a dynamic system, it follows that information is created all the time in connection with the emergence of biological form. In mathematical terms creation of biological form is a recursive mapping of form transitions based on successive information generation. In this context DNA is what enables life to play aikido[2] with mathematics, physics and chemistry and build order upon order in an unsurpassed way.

Open physical systems are sustained non-equilibrium systems exchanging matter and energy with the environment. All living systems are certainly open physical systems, but so are very many inanimate systems that together span immense spatial and temporal scales. The latter group also shows intriguing morphogenetic capacity due to self-organization and emergence (Cross and Greenside, 2009). We are starting to understand some of the mechanisms underlying this capacity (Buka and Kramer, 1996; Hoyle, 2006; Cross and Greenside, 2009; Desai and Kapral, 2009) and how to exploit this knowledge in technology development (Buka and Kramer, 1996; Desai and Kapral, 2009). Living systems do not have exclusive ownership to phenomena like self-assembly, self-organization, emergence, two-way causation between lower- and higher-level system dynamics features, and order creation through local reduction of entropy.

The form-generating capacity of living systems dwarfs that of inanimate open systems, however. I think the major reason for this is that the presence of DNA allows a system to induce perturbations of its own dynamics as a function of its own system state or phenome (Fig. 1). This feature, which should not be equated with two-way causality, enables living systems to create order upon order and attain forms in *morphospace* that are beyond reach for any open physical system that relies on the information generation that follows from the autonomous unfolding of the system per se. In mathematical terms, DNA allows successive *within-system* state space changes both in terms of state space trajectory changes and reconfiguration of the state space as such. In contrast to non-living open physical systems, the presence of DNA is a systems-structure that *capacitates living systems to self-transcend* – not beyond the dictums of physics and chemistry, but beyond those morphogenetic limits that exist for non-living open physical systems in general. The term 'transcendent' is defined as something going beyond or exceeding usual limits, and I use the term *self-transcendence* deliberately because here the transcendence is a function of system state.

It should be noted that according to the above scheme there is no direct causal arrow from genotype to phenotype in the sense that DNA is responsible for exerting a direct effect as *a sub-system* on the system dynamics. The causality flows from the system state through a *change in use* of DNA (as an inert system component) that results in a change in the production of RNA and protein, that in

[2] A Japanese martial art that is performed by blending with the motion of the attacker and redirecting the attacker's momentum rather than opposing it directly.