TOPICAL REVIEW — Statistical physics and complex systems

Equivalent formulations of "the equation of life"*

Ao Ping(敖 平)†

Shanghai Center for Systems Biomedicine, Key Laboratory of Systems Biomedicine of Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, *China*

(Received 19 March 2014; published online 22 May 2014)

Motivated by progress in theoretical biology a recent proposal on a general and quantitative dynamical framework for nonequilibrium processes and dynamics of complex systems is briefly reviewed. It is nothing but the evolutionary process discovered by Charles Darwin and Alfred Wallace. Such general and structured dynamics may be tentatively named "the equation of life". Three equivalent formulations are discussed, and it is also pointed out that such a quantitative dynamical framework leads naturally to the powerful Boltzmann–Gibbs distribution and the second law in physics. In this way, the equation of life provides a logically consistent foundation for thermodynamics. This view clarifies a particular outstanding problem and further suggests a unifying principle for physics and biology.

Keywords: Darwin, evolution, nonequilibrum processes, stochastic dynamics

PACS: 05.70.Ln, 05.10.Gg, 72.70.+m, 87.15.Ya **DOI:** 10.1088/1674-1056/23/7/070513

1. Introduction

Let us start with a disclaimer to avoid any impression of dogma that might be inferred from our title: neither the equation nor the universal law proposed here implies predetermined or unchangeable rules. Rather, our proposal expresses the fact that within all of the existing empirical evidence, there is no valid counter example against such a law, in much the same manner, say, as the situations for special relativity and quantum mechanics in physics. In addition, there is no viable indication that it could be shown to be violated in the foreseeable future. Given this qualification, or precisely because of it, many biologists would argue that there can be no universal law in biology: the biological phenomena are too diverse to permit a unifying framework. We emphatically disagree with such a view and wish to reason here that the principle of natural selection discovered by Charles Darwin and Alfred Wallace is a candidate. Because the existence of a general quantitative framework has been questioned on one side[1,2] as well as longed for from the other side,[3] in the following we argue it from three different perspectives, extending an early reasoning.[4]

2. First formulation: Universal algorithm

We immediately face a thorny question: is it known that there is not even a consensus among biologists on what the principle of natural selection would mean?[1] We recognize the implication of such a question and we still favor the existence of a unifying framework. The crucial point is the recognition of the equal importance of variation and selection, that is, we propose, admittedly not the first to do so, that the principle of natural selection should be interpreted by the following word equation:[5,6]

$$\text{Evolution by Variation and Selection.} \tag{1}$$

It is known that any partial emphasis on variation or selection has led to controversy, namely, take the notable example of neutral theories vs selection theories. The word equation above immediately suggests itself as a probabilistic updating rule on successive processes. It further suggests that the evolutionary dynamical structure discovered by Darwin and Wallace is actually a general algorithm from the modern computational point of view.[1,7] In fact, it is manifestly universal because it does not specify how it would be realized. It is not even necessarily confined to biology: all algorithms have such generality. Thus, evolution by variation and selection can indeed imply a universal law.

There are, however, two main concerns which indicate that such an algorithmic interpretation is not completely satisfactory. Alternative formulations are needed. We should hasten to add before moving on that there is nothing wrong with the universal algorithm understanding of evolutionary dynamics; in many cases, such understanding can even be advantageous. The first concern is of a pragmatic nature. Suppose for a given evolutionary dynamics we already know the explicit algorithm and wish to know the possible outcomes that follow from a given initial condition. From the algorithmic point of view, the only way we can find them out is to run the algorithm, either on digital or analog computers. The latter could be real implementations in biology by appropriate biological processes. Because of the stochastic nature of such algorithms, many repeated runs would be also needed to decide which one

*Project supported in part by the 985 Project of Shanghai Jiao Tong University, China.

†Corresponding author. E-mail: aoping@sjtu.edu.cn

© 2014 Chinese Physical Society and IOP Publishing Ltd

would be more likely to occur. The problem is that such an algorithmic approach could be very costly, and/or would easily take an astronomically long time. Hence, we are often forced to look for alternative approaches that can lead to the needed information. The second concern is of a human nature: even if a given problem can be nicely handled in algorithmic manner, we wish to understand the working of such processes by ourselves and will not be satisfied to let computers do the task in an input-output, black box manner. For example, we may want to understand the "fitness" and associated propensity debates even since Darwin and Wallace.[1] Though it is in principle possible to exercise the algorithmic view, it has not given us very much so far.

3. Second formulation: Dynamical structure

During past 100 years, ceaseless efforts have been made to understand the quantitative theoretical structure embedded in the dynamical theory of Darwin and Wallace. Two key concepts have emerged. One is the fundamental theorem of natural selection (FTNS), which stresses the importance of variation, proposed by Ronald Fisher.[8] Another is the adaptive landscape proposed by Sewall Wright,[9] which captures the essence of "fitness" and the similar concept of developmental process originated by Conrad Waddington.[10] Nevertheless, even those two important concepts have remained controversial since being proposed in the early 1930s. Fortunately, with the recent progress in biological modeling and associated mathematical formulations, those controversies have been resolved.[5,11] The algorithmic form of evolution by variation and selection has indeed been transformed into a set of special form mathematical equations. The adaptive landscape is represented by a potential function. The adaptive and nonconservative dynamics are associated with variation, and there is a conservative dynamics. The last part has not been emphasized in the field of biology, though various authors have implicitly noted its existence. This recent result has been summarized into a concise mathematical equation[5,6]

$$[\boldsymbol{A}+\boldsymbol{T}]\partial_t q = \partial_q \psi + \xi(q,t). \tag{2}$$

Here the potential function ψ quantifies the adaptive landscape, and the relation between the ascendant matrix $\boldsymbol{A}$ and the fluctuation ξ generalizes Fisher's FTNS.[5] Within the new formulation of evolutionary dynamics, one can show that, from a logical point of view, the formulation is completely consistent with the well-known conservative dynamics in the physical sciences. More importantly, it gives rise naturally to the Boltzmann–Gibbs distribution and thermodynamics, and hence the second law of thermodynamics.[6] There have been many confused and failed attempts to "derive" the evolutionary dynamics from thermodynamics. We now know the reason: the logical order in those failed efforts was reversed. Thus, it becomes clear that evolution by variation and selection is indeed a universal dynamical law. This certainly corroborates well with the successful applications of Boltzmann–Gibbs distribution in molecular biology[12] and beyond.

To careful readers, two conceptual concerns may emerge. First, the universal law appears to be too universal. No alternative dynamical law could be conceived, at least by us, if apparent *ad hoc* assumptions are excluded. It appears to be the only game in the town. While it confirms the generality of such a dynamical law, it may remind us of the notorious tautology argument of Karl Popper. We admit our lack of imagination to conceive an alternative, but do not think there is any tautology in the present formulation. Perhaps we have to acknowledge that our world has been structured in this way.

Second, this dynamical law could not be the whole description for concrete evolutionary processes. Indeed, it is not. From the algorithmic point of view, it must have a real materialistic substrate to implement the algorithm, being mechanical instruments, vacuum tubes, silicon transistors, or something else. Biologically, it is known that evolution dynamics have two major parts, that is, two kinds of laws: one is on the universal dynamical structure, already discussed; the other is on how each of the dynamical elements would be realized. For example, for the above word equation, three things have to be specified in order to compare it with a given empirical process: what would be the successive process, such as sexual, asexual, or others? How would variations arise, intrinsic ones such as generated by the organisms themselves or extrinsic ones completely controlled by environments? What would dominate the selection processes, namely, the innate structure of the organism, the structure of environments, or the combination of both? This amounts to specifying related quantities in above mathematical equation: the choice of the proper state variable q, the symmetric matrix $\boldsymbol{A}$ representing adaptation, the antisymmetric matrix $\boldsymbol{T}$ representing the conservative aspect of dynamics, and the potential function ψ representing the adaptive landscape. The noise would be automatically related to the symmetric matrix via the F-theorem, the generalization of the fundamental theorem of natural selection. Historically, while Darwin fully understood the universal dynamics discovered by himself and Wallace, he was also keenly aware of the absence of the second type of universal law: the hereditary implementation of evolution. He tried and failed to uncover such a law. In hindsight, it is clear what he missed: the gene/DNA structure of Gregor Mendel, James Watson, and Francis Crick.

At this point, readers with preference in holistic theories in biology may comment that we are retreating from that goal by dividing biological laws and thus general descriptions of biological processes into two kinds, the dynamical laws, and the laws that implement the dynamics. Indeed, in our view, there is as yet no such universal theory that could put everything

known in biology together, something similar to Einstein's general relativity, which profoundly integrates special relativity and gravity. This is not necessarily a drawback, given our present incomplete understanding of hierarchical structures in biology. The classification of biological laws into two major kinds has, on the other hand, two positive consequences. First, it is clear by the demonstration above that evolutionary dynamics is indeed universal, applicable to the whole of biology. Taking the favorable example of self-organization into holistic consideration as an illustration, the dynamics of evolution by variation and selection clearly occurs before, during and after self-organization. Second, by admitting the incompleteness of our current knowledge, we show confidence in our successors: among them there may be future Darwins, Wallaces, or Einsteins who may reach better syntheses based on our efforts, but deeper than what we have reached. In summary, for the recently posted question on the existence of "a theory of biological relativity",[3] the answer here is both yes and no: from the purely dynamical perspective of special relativity, it does exist, and here we do have a universal dynamical equation for evolution; while for the unifying theory of general relativity of dynamics and gravity, it does not.

One may wonder about the use of such a universal dynamical law. Let us list two. The theoretical structure of universal dynamics can be used to formulate biological problems in a precise mathematical language. Many features that are vague in the current verbal formulation can then become evident. For example, the potential function corresponding to the adaptive landscape provides a natural quantification for robustness and stability.[13,14] The historical perspective embedded in evolutionary dynamics can be used to reveal the working of a given biological process, another fact in biology. For example, because of the common ancestry of organisms, a discovery made about one organism may lend crucial hints for another, such as in cancer research.[15,16] Other utilities of such general theory are also discussed by many others.[3]

4. Third formulation: Entangled selection and variation

Our reasoning above may be summarized in plain language. The following is what the universal evolutionary dynamics has naturally suggested us:

with selection, complete disorder is impossible;
with variation, complete order is also impossible. (3)

The first half of the statement is on the creative nature of variation and disorder: order can grow out of them in the presence of selection. The second part is, for readers well versed in thermodynamics, a restatement of the second order of thermodynamics, which indicates that constant effort is needed to maintain order. Under the operation of such dynamics, life may sooner or later appear somewhere. Thus, such an equation, (1), (2), or (3) can be truly called the equation of life.

5. Perspective

We have summarized reasons that the evolutionary dynamics discovered by Darwin and Wallace can be formulated into a quantitative mathematical framework with three equivalent formulations: the traditional form of Darwin and Wallace, the stochastic dynamic structural formulation, which extends the concepts of Fisher's FTNS and Wright's adaptive landscape, and the negative language formulation. This theoretical framework is apparently valid to all levels in biology. It may stand as a viable candidate to satisfy the requirement for a theory of biological relativity after the human genome project era. It is possible that such formulation may generate applications in related biological fields[17] and new connection to other fields.[18] It may further stimulate interest in mathematical structures embedded in stochastic processes.[19–22]

Acknowledgments

Critical and constructive discussion with various people, especially David Galas, Manyuan Long, and Hong Qian, were very helpful in the long process to distill the ideas presented here.

References

[1] Hull D L and Ruse M 2007 *The Cambridge Companion to the Philosophy of Biology* (Cambridge: Cambridge University Press)
[2] Bock W J 2007 *J. Zool. Syst. Evol. Res.* **45** 89
[3] Noble D 2010 *Phil. Trans. R. Soc. A* **368** 1125
[4] Ao P 2008 *Bioinformatics and Biology Insights* **2** 201
[5] Ao P 2005 *Phys. Life Rev.* **2** 117
[6] Ao P 2008 *Commun. Theor. Phys.* **49** 1073
[7] Dennett D C 1995 *Darwin's Dangerous Idea: Evolution and the Meanings of Life* (New York: Simon and Schuster)
[8] Fisher R A 1930 *The Genetical Theory of Natural Selection* (Oxford: Clarendon)
[9] Wright S 1932 *The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution, in Proceedings of the Sixth International Congress of Genetics* (D. F. Jones ed.) (Menasha, WI: Brooklyn Botanic Garden) pp. 356–366
[10] Waddington C H 1957 *The Strategy of the Genes: A Discussion of Some Aspect of Theoretical Biology* (New York: The MacMillan Company)
[11] Ao P 2009 *J. Genet. Genomics 36 63*
[12] Dubois J M, Ouanounou G and Rouzaire-Dubois B 2009 *Prog. Biophys. Mol. Bio.* **99** 87
[13] Zhu X M, Yin L, Hood L and Ao P 2004 *J. Bioinf. Comput. Biol.* **2** 785
[14] Wang G W, Zhu X M, Gu J R and Ao P 2014 *Interface Focus* **4** 20130064
[15] Weinberg R A 2007 *The Biology of Cancer* (New York: Taylor and Francis)
[16] Ao P, Galas D, Hood L and Zhu X M 2008 *Medical Hypotheses* **70** 678
[17] Friston K and Ao P 2012 *Computational and Mathematical Methods in Medicine* **2012** 937860
[18] Qian H 2013 *Quantitative Biology* **1** 50
[19] Ge H and Qian H 2012 *Int. J. Mod. Phys. B* **26** 1230012
[20] Yuan R S and Ao P 2012 *J. Stat. Mech.* **2012** P07010
[21] Ao P, Chen T Q and Shi J H 2013 *Chin. Phys. Lett.* **30** 070201
[22] Qian H 2013 *J. Math. Phys.* **54** 053302