



ELSEVIER

Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com



Biological modelling/Biomodélisation

A new formal perspective on ‘Cambrian explosions’



Rodrick Wallace

Division of Epidemiology, The New York State Psychiatric Institute, Box 47, NSPI, 1051 Riverside Drive, New York, NY, 10032, USA

ARTICLE INFO

Article history:

Received 10 October 2013

Accepted after revision 4 November 2013

Available online 4 December 2013

Keywords:

Evolution

Information theory

Large deviations

Phase transition

Punctuation

ABSTRACT

The ‘Cambrian explosion’ 500 Myr ago saw a relatively sudden proliferation of organism Bauplan and ecosystem niche structure that continues to haunt evolutionary biology. Here, adapting standard methods from information theory and statistical mechanics, we model the phenomenon as a noise-driven phase transition, in the context of deep-time relaxation of current path-dependent evolutionary constraints. The result is analogous to recent suggestions that multiple ‘explosions’ of increasing complexity in the genetic code were driven by rising intensities of available metabolic free energy. In the absence of severe path-dependent lock-in, ‘Cambrian explosions’ are standard features of blind evolutionary process, representing outliers in the ongoing routine of evolutionary punctuated equilibrium.

© 2013 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

Over the years, evolutionary scientists have explored in detail the ‘Cambrian explosion’ of half a billion years ago, a punctuated change in the diversity of life (e.g., [1–4]) sometimes seen in popular and religious literature as challenging evolutionary theory. Here we demonstrate, from a somewhat novel perspective, that a relatively modest formal exercise in that theory accounts neatly for such ‘explosions’ early on, before path-dependent lock-in of essential biochemical, gene regulation, and more generally biological, Bauplans. The approach illuminates as well the current rapid evolution of viral/viroid species and quasi-species [5]. The underlying mechanisms appear much the same. Similar discussions, under the rubric ‘punctuated equilibrium’, have, of course, long been in the literature (e.g., Gould [2], and references therein). The work here supports that view, but provides a new theoretical line of argument.

The approach follows that of Wallace [6], where it is argued multiple punctuated ecosystem regime changes in metabolic free energy broadly similar to the aerobic

transition enabled a punctuated sequence of increasingly complex genetic codes and protein translators. Then, in a manner similar to the serial endosymbiosis effecting the eukaryotic transition, codes and translators coevolved until the ancestor of the present narrow spectrum of protein machineries became locked-in by evolutionary path dependence at a relatively modest level of fitness reflecting a modest embedding metabolic free energy ecology [7]. The search for such preaerobic biochemical Cambrian-like ‘explosions’ is, of course, much hampered by the absence of early chemical evolution from currently-studied fossil records.

Population genetics defines evolution by changes in allele frequencies [8,9]. Evolutionary game dynamics track such shifts under natural selection using the replicator model of Taylor and Jonker [10]. These and related mathematical models purport to be both a necessary and sufficient definition of evolution across disciplines from biology to economics, albeit with sometimes scathing dissent (e.g., Roca et al. [11]).

Wallace [6,12–16], in contrast, proposes a set of necessary conditions statistical models extending evolutionary theory via the asymptotic limit theorems of communication theory. The method represents genetic heritage, regulated gene expression, and the surrounding

Email address: rodrick.wallace@gmail.com.

environment as interacting information sources. A fundamental insight is that gene expression can be directly seen as a cognitive phenomenon associated with a ‘dual’ information source, while the embedding environment’s systematic regularities ‘remember’ imposed changes, resulting in a coevolutionary process in the sense of Champagnat et al. [17] that is recorded jointly in, genes, gene expression regulation, and the embedding environment. See references [6,12–16,18] for details.

The focus here is on the effect of ‘large deviations’ representing transitions between the quasi-stable modes of such systems that are analogous to game-theoretic Evolutionary Stable Strategies. Evolutionary path dependence, in general, limits such possible excursions to high-probability sequences consistent with, if not originating in, previous evolutionary trajectories: after some three billion years, however, most multicellular organisms evolve, they retain their basic Bauplan, with only relatively small non-fatal variations currently allowed.

We are interested in matters half a billion years ago, before path dependence solidly locked-in possible large deviations excursions.

In essence, a sufficiently large number of allowed large deviation trajectories leads, consequently, to many available quasi-equilibrium states. These, in turn, can be treated as an ensemble, i.e., in a manner similar to the statistical mechanics perspective on critical phenomena. This allows a new approach to rapid evolutionary change – in deep time for multicellular organisms, and in real time for current populations of viruses and viroids.

That is, even today, while incorporation of long-term path dependence drastically reduces possible evolutionary dynamics in higher organisms, viral or viroid evolution can be explored in the same way, driven by ‘noise’ defined as much by policy and socioeconomic structure as by reassortment and generation time [e.g., 5].

2. The basic model

Following Wallace and Wallace [12,13], assume there are n populations interacting with an embedding environment represented by an information source Z . The genetic and (cognitive) gene expression processes associated with each species i are represented as information sources X_i, Y_i respectively. These information sources undergo a ‘coevolutionary’ interaction in the sense of Champagnat et al. [17], producing a joint information source uncertainty [19] for the full system as

$$H(X_1, Y_1, \dots, X_n, Y_n, Z) \quad (1)$$

In addition, Feynman’s [20] insight that information is a form of free energy allows definition of an entropy-analog as

$$S \equiv H - Q_j \sum_j \partial H / \partial Q_j \quad (2)$$

The Q_i are taken as driving parameters that may include, but are not limited to, the Shannon uncertainties of the underlying information sources. See Cover and Thomas [19] for a basic introduction to information theory.

Again, in the spirit of Champagnat et al. [17], we can characterize the dynamics of the system in terms of Onsager-like non-equilibrium thermodynamics in the gradients of S as the set of stochastic differential equations [21],

$$dQ_t^i = L_i \left(\partial S / \partial Q^1 \dots \partial S / \partial Q^m, t \right) dt + \sum_k \sigma_k^i \left(\partial S / \partial Q^1 \dots \partial S / \partial Q^m, t \right) dB_k, \quad (3)$$

where the B_k represent noise terms having particular forms of quadratic variation. See standard references on stochastic differential equations for details.

This can be more simply written as:

$$dQ_t^i = L_i(\mathbf{Q}, t) dt + \sum_k \sigma_k^i(\mathbf{Q}, t) dB_k, \quad (4)$$

where $\mathbf{Q} \equiv (Q^1, \dots, Q^m)$.

Following the arguments of Champagnat et al. [17], this is a coevolutionary structure, where fundamental dynamics are determined by component interactions:

- setting the expectation of Eq. (4) equal to zero and solving for stationary points gives attractor states since the noise terms preclude unstable equilibria. These are analogous to the evolutionarily stable states of evolutionary game theory;
- this system may, however, converge to limit cycle or pseudorandom ‘strange attractor’ behaviors similar to thrashing in which the system seems to chase its tail endlessly within a limited venue – the ‘Red Queen’;
- what is ‘converged’ to in any case is not a simple state or limit cycle of states. Rather it is an equivalence class, or set of them, of highly dynamic information sources coupled by mutual interaction through crosstalk and other interactions. Thus ‘stability’ in this structure represents particular patterns of ongoing dynamics rather than some identifiable static configuration;
- applying Ito’s chain rule for stochastic differential equations to the (Q_j^i) and taking expectations allows calculation of variances. These may depend very powerfully on a system’s defining structural constants, leading to significant instabilities [22], something we will explore more fully below.

3. Large deviations: iterating the model

As Champagnat et al. [17] note, shifts between the quasi-equilibria of a coevolutionary system can be addressed by the large deviations formalism. The dynamics of drift away from trajectories predicted by the canonical equation can be investigated by considering the asymptotic of the probability of ‘rare events’ for the sample paths of the diffusion.

‘Rare events’ are the diffusion paths drifting far away from the direct solutions of the canonical equation. The probability of such rare events is governed by a large deviation principle, driven by a ‘rate function’ \mathbf{Iscr} ; that can be expressed in terms of the parameters of the diffusion.

This result can be used to study long-time behavior of the diffusion process when there are multiple attractive

singularities. Under proper conditions, the most likely path followed by the diffusion when exiting a basin of attraction is the one minimizing the rate function **Iscr**; over all the appropriate trajectories.

An essential fact of large deviations theory, however, is that the rate function **Iscr**; almost always has the canonical form

$$\mathbf{Iscr}; = - \sum_j \mathbf{P}_j \log(\mathbf{P}_j) \tag{5}$$

for some probability distribution, i.e., the uncertainty of an information source [19]. This result goes under a number of names; Sanov’s Theorem, Cramer’s Theorem, the Gartner-Ellis Theorem, the Shannon-McMillan Theorem, etc. [23].

The argument directly complements Eq. (4), now seen as subject to large deviations that can themselves be described as the output of an information source L_D defining **Iscr**;, driving or defining Q^i -parameters that can trigger punctuated shifts between quasi-stable system modes.

This is now a common perspective in systems biology (e.g., Kitano [24]).

Not all large deviations are possible: only those consistent with the high-probability paths defined by the information source L_D will take place.

Recall from the Shannon-McMillan Theorem [25] that the output streams of an information source can be divided into two sets, one very large that represents nonsense statements of vanishingly small probability, and one very small of high probability representing those statements consistent with the inherent ‘grammar’ and ‘syntax’ of the information source. Again, whatever higher-order multi-cellular evolution takes place, some equivalent of backbone and blood remains.

Thus we could now rewrite Eq. (1) as:

$$H_L(X_1, Y_1, \dots, X_n, Y_n, Z, L_D), \tag{6}$$

where we have explicitly incorporated the ‘large deviations’ information source L_D that defines high probability evolutionary excursions for this system.

Again carrying out the argument leading to Eq. (4), we arrive at another set of quasi-stable modes, but possibly very much changed in number; either branched outward in time by a wave of speciation, or decreased through a wave of extinction. Iterating the models backwards in time constitutes a cladistic or coalescent analysis.

4. Extinction

A simple extinction model leads to significant extension of the theory.

Let $N_t \geq 0$ represent the number of individuals of a particular species at time t . The simplest dynamic model, in this formulation, is then something like

$$dN_t = -\alpha N_t |N_t - N_c| dt + \sigma N_t dW_t, \tag{7}$$

where N_c is the ecological carrying capacity for the species, α is a characteristic time constant, σ is a ‘noise’ index, and dW_t represents white noise.

Taking the expectation of Eq. (7), the possible equilibrium values of N_t are either zero or N_c . Applying the Ito chain rule [26] to the second moment in N_t , i.e., to N_t^2 , a somewhat lengthy calculation finds there can be no real second moment unless

$$\sigma^2 < 2\alpha N_c \tag{8}$$

That is, unless Eq. (8) holds – the product of the rate of population change and carrying capacity is sufficiently large – noise-driven fluctuations will inevitably drive the species to extinction.

A similar SDE approach has been used to model noise-driven criticality in physical systems [27–29], suggesting that a more conventional phase transition methodology may provide particular insight.

5. Evolution under relaxed path-dependence

In general, for current higher plants and animals, the number of quasi-equilibria available to the system defined by Eq. (4), or to its generalization via Eq. (6), will be relatively small, a consequence of long-term lock-in by path-dependent evolutionary process. The same cannot be said, however, for virus/viroid species or quasi-species, to which can be applied more general methods (e.g., Wallace and Wallace [5]) that may also represent key processes acting half a billion years in the past.

Under such a relaxation assumption, the speciation/extinction large deviations information source L_D is far less constrained, and there will be very many possible quasi-stable states available for transition, analogous to an ensemble in statistical mechanics.

The noise parameter in Eq. (7) can, from the arguments of the previous section, then be interpreted as a kind of temperature-analog, and N_t as an order parameter that, like magnetization or ice crystal form, vanishes above a critical value of σ . This leads – in particular for something as protean as an influenza virus or an HIV quasi-species – to a relatively simple statistical mechanics analog built on the H_L of Eq. (6).

Define a pseudoprobability for quasi-stable mode j as:

$$P_j = \frac{\exp[-H_L^j/\kappa\sigma]}{\sum_i \exp[-H_L^i/\kappa\sigma]}, \tag{9}$$

where κ is a scaling constant and σ is a noise intensity.

Next, define a Morse function F , in the sense used by Pettini [30], as:

$$\exp[-F/\kappa\sigma] \equiv \sum_i \exp[-H_L^i/\kappa\sigma] \tag{10}$$

Apply Pettini’s topological hypothesis to F , taking N_j , the number of members of species (or quasi-species) j as a kind of ‘order parameter’, in Landau’s sense [31]. Then σ is seen as a very general temperature-like measure whose changes drive punctuated topological alterations in the underlying ecological and coevolutionary structures associated with the Morse function F .

Such topological changes, following Pettini’s arguments, can be far more general than indexed by the

simple Landau-type critical point phase transition in an order parameter.

Indeed, the results of Wallace [6], regarding the complexity of the genetic code, could be directly reframed in terms of available metabolic free energy intensity leading to something like Eqs. (9) and (10). Then the term $\kappa\sigma$ is replaced by κM , where M is a measure of metabolic free energy intensity, and the H_L^j represent the Shannon uncertainties in the transmission of information between codon machinery and amino acid machinery.

Increasing M then leads to more complex genetic codes, i.e., those having higher measures of symmetry, in the Landau sense, until evolutionary lock-in took place at a relatively low level of coding efficiency. Canfield et al. [7], in their Tables 1 and 2, provide a range of possible electron donor and acceptor mechanisms that may have been available to fuel metabolic free energy under preaerobic conditions, in the context of anoxygenic phototrophs. They speculate that the most active early ecosystems were probably driven by the cycling of H_2 and Fe^{2+} , providing relatively low free energy intensities for metabolic process.

6. Discussion and conclusions

The topological changes inherent in a relaxed path-dependence model can represent a great variety of fundamental and highly punctuated coevolutionary alterations, since the underlying species and quasi-species are not so sharply limited by path-dependent evolutionary trajectory in the manner that constrains variation in most current higher organisms.

That is, under an assumption of less lock-in constraint a half-billion years ago, such a model accounts well for the observed Cambrian Bauplan explosion, and possibly as well for much earlier multiple ‘explosions’ in genetic codes postulated by Wallace [6].

In general, according to our model, the degree of punctuation in evolutionary punctuated equilibrium [2] will depend strongly on the richness of the distribution of quasi-equilibria to be associated with Eq. (4). ‘Cambrian explosions’ therefore require a dense statistical ensemble of them, our analog to the ‘roughening of the fitness landscape’ described as necessary in Marshall [3].

The ‘noise’, in the Cambrian case, may have been a kind of species or quasi-species isolation, via separation by geographic or ecological niche, that was (relatively) suddenly lifted. Permitting greater interaction – lowering σ – may then have triggered a long series of rapid coevolutionary transitions, an inverse, as it were, of the observations of Hanski et al. [32] on species extinction via habitat fragmentation.

As Marshall put it [3],

“With the advent of ecological interactions between macroscopic adults... especially... predation..., the number of needs each organism had to meet must have increased markedly: Now there were myriad predators to contend with, and a myriad number of ways to avoid them, which in turn led to more specialized ways of predation as different species developed different avoidance strategies, etc. The combinatoric richness

already present in the Ediacaran genome was extracted through the richness of biotic interaction as the Cambrian ‘explosion’ unfolded...”

Other dynamics may also have contributed. Von Bloh et al. [33], for example, develop a mathematical model of the Cambrian explosion as explicitly related to rapid cooling, which amplified the spread of complex multicellular life via ‘nonlinear geosphere-biosphere interactions’. Cooling events, in their model, could trigger transition from a quasi-equilibrium without, to one with, complex multicellular life. This involved a positive feedback between the spread of biosphere, increased silicate weathering, and a consequent cooling of the climate.

In any event, 20–80 million year duration of the Cambrian ‘explosion’ provides ample time for a broad-scale evolutionary divergence in multicellular life forms under conditions of relaxed path dependence. Similar explosive divergences in early evolution of the genetic code may also have occurred.

The general inference is that, in the absence of severe path-dependent lock-in, ‘Cambrian explosions’ can be a common feature of blind evolutionary process, representing expected outliers in the ongoing routine of evolutionary punctuated equilibrium [2].

Disclosure of interest

The author declares that he has no conflicts of interest concerning this article.

References

- [1] D. Erwin, J. Valentine, *The Cambrian Explosion: The construction of animal biodiversity*, Roberts and Company, Greenwood Village, 2013.
- [2] S. Gould, *The Structure of Evolutionary Theory*, Harvard University Press, Cambridge, MA, 2002.
- [3] C. Marshall, Explaining the Cambrian ‘explosion’ of animals, *Annu. Rev. Earth Planet. Sci.* 34 (2006) 355–384.
- [4] H. Whittington, *The Burgess Shale*, Yale University Press, New Haven, CT, 1985.
- [5] R. Wallace, R.G. Wallace, Blowback: new formal perspectives on agri-business-driven pathogen evolution. Submitted.
- [6] R. Wallace, Metabolic constraints on the evolution of genetic codes: did multiple ‘preaerobic’ ecosystem transitions entrain richer dialects via serial endosymbiosis? *Trans. Comput. Syst. Biol.* XIV, LNBI 7625 (2012) 204–232.
- [7] D. Canfield, M. Rosing, C. Bjerrum, Early anaerobic metabolisms, *Philos. Trans. Roy. Soc. B* 361 (2006) 1819–1836.
- [8] W. Ewens, *Mathematical Population Genetics*, Springer, New York, 2004.
- [9] D. Hartl, A. Clark, *Principles of Population Genetics*, Sinaur Associates, Sunderland, MA, 2006.
- [10] P. Taylor, L. Jonker, Evolutionary stable strategies and game dynamics, *Math. Biosci.* 40 (1978) 145–156.
- [11] C. Roca, J. Suesta, A. Sanchez, Evolutionary game theory: temporal and spatial effects beyond replicator dynamics, *Phys. Life Rev.* 6 (2009) 208–249.
- [12] R. Wallace, Expanding the modern synthesis, *C. R. Biologies* 334 (2010) 263–268.
- [13] R. Wallace, A formal approach to evolution as self-referential language, *BioSystems* 106 (2011) 36–44.
- [14] R. Wallace, Consciousness, crosstalk, and the mereological fallacy: an evolutionary perspective, *Phys. Life Rev.* 9 (2012) 426–453.
- [15] R. Wallace, A new formal approach to evolutionary processes in socio-economic systems, *J. Evol. Econ.* 23 (2013) 1–15.
- [16] R. Wallace, Cognition and biology: perspectives from information theory, *Cogn. Process.* (2013), <http://dx.doi.org/10.1007/s10339-013-0573-1>.

- [17] N. Champagnat, R. Ferriere, S. Meleard, Unifying evolutionary dynamics: from individual stochastic process to macroscopic models, *Theor. Popul. Biol.* 69 (2006) 297–321.
- [18] R. Wallace, D. Wallace, Punctuated equilibrium in statistical models of generalized coevolutionary resilience: how sudden ecosystem transitions can entrain both phenotype expression and Darwinian selection, *Trans. Comput. Syst. Biol.* IX, LNBI 5121 (2008) 23–85.
- [19] T. Cover, J. Thomas, *Elements of Information Theory*, second ed., Wiley, New York, 2006.
- [20] R. Feynman, *Feynman Lectures on Computation*, Westview Press, 2000.
- [21] S. de Groot, P. Mazur, *Non-Equilibrium Thermodynamics*, Dover, New York, 1984.
- [22] R. Khasminskii, *Stochastic Stability of Differential Equations*, second ed., Springer, New York, 2012.
- [23] A. Dembo, O. Zeitouni, *Large Deviations: Techniques and Applications*, Springer, New York, 1998.
- [24] H. Kitano, Biological robustness, *Nat. Genet.* 5 (2004) 826–837.
- [25] A. Khinchin, *Mathematical Foundations of Information Theory*, Dover Publications, New York, 1957.
- [26] P. Protter, *Stochastic Integration and Differential Equations*, Springer, New York, 1990.
- [27] W. Horsthemke, R. Lefever, *Noise-induced Transitions, Theory and Applications in Physics, Chemistry, and Biology*, vol. 15, Springer, New York, 2006.
- [28] C. Van den Broeck, J. Parrondo, R. Toral, Noise-induced nonequilibrium phase transition, *Phys. Rev. Lett.* 73 (1994) 3395–3398.
- [29] C. Van den Broeck, J. Parrondo, R. Toral, R. Kawai, Nonequilibrium phase transitions induced by multiplicative noise, *Phys. Rev. E* 55 (1997) 4084–4094.
- [30] M. Pettini, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Springer, New York, 2007.
- [31] L. Landau, E. Lifshitz, *Statistical Physics. Part I*, third ed., Elsevier, New York, 2007.
- [32] I. Hanski, G. Zurita, M.I. Bellocq, J. Rybicki, Species-fragmented area relationship, *Proc. Natl. Acad. Sci. USA* (2013), <http://dx.doi.org/10.1073/pnas.1311491110>.
- [33] W. von Bloh, C. Bounama, S. Franck, Cambrian explosion triggered by geosphere-biosphere feedbacks, *Geophys. Res. Lett.* 30 (2003) 1963–1968.