

1
2

3 Article

4 Looking for LUCA

5 Minna Koskela ¹ and Arto Annala ^{1,2,3,*}

6 ¹ Department of Biosciences, Viikinkaari 1, FI-00014 University of Helsinki, Finland

7 ² Institute of Biotechnology, Viikinkaari 1, FI-00014 University of Helsinki, Finland

8 ³ Department of Physics, Gustaf Hällströmin katu 2, FI-00014 University of Helsinki, Finland

9 E-Mails: minna.m.koskela@helsinki.fi (M. K.); arto.annala@helsinki.fi (A. A.)

10 * Author to whom correspondence should be addressed; Tel.: +358-9-191-50629; Fax: +358-9-191-
11 50639

12 Received: 26.11.2011 / Revised: 19.12.2011 / Accepted: / Published:
13

14 **Abstract:** Genomic sequences across diverse species seem to align towards a common
15 ancestry, eventually implying that eons ago some universal antecedent organism would have
16 lived on the face of Earth. However, when evolution is understood not only as a biological
17 process but as a general thermodynamic process, it will become apparent that the quest for
18 the last universal common ancestor is an unattainable attempt. Ambiguities in alignments
19 are unavoidable because the driving forces and paths of evolution cannot be separated from
20 each other. Thus tracking down life's origin is by its nature a non-computable task. The
21 thermodynamic tenet clarifies that evolution is a path-dependent process of least-time
22 consumption of free energy. The natural process is without a demarcation line between
23 animate and inanimate.

24 **Keywords:** evolution; geodesic; intractability; sequence alignment; the principle of least
25 action; universal ancestor
26

27 1. Introduction

28 The quest for the last universal common ancestor (LUCA) roots back to the much-quoted closing
29 paragraph of On the Origin of Species where Charles Darwin infers: "...that probably all the organic
30 beings which have ever lived on this earth have descended from some one primordial form, into which

1 life was first breathed.” The notion of common origin is even more motivated today since biochemical,
2 molecular biology and genetic studies have disclosed that apart from similarities in phenotypes, many
3 molecular characteristics of life are also universal. Most notably, the same metabolites, proteins and
4 nucleic acids (including their specific handedness) are found throughout nature [1]. Also, the genetic
5 code is next to identical for all life forms [2]. Most importantly, similarities in genomic sequences
6 across life’s major branches imply universal common ancestry [3]. This conclusion about the common
7 origin seems not to be obscured by lateral gene transfer [4] or by convergent evolution to sequence
8 similarity [5]. Even if life had emerged multiple times, LUCA has been regarded as an outcome of
9 convergent evolution, e.g., as a single recombining population of early organisms [6] or as the only
10 one species that survived in the course of early evolution [7]. The probability of common ancestry has
11 been considered by statistical tests to be substantially higher than that of multi-ancestor models [8].
12 However, likelihood analyses may be illusory in proving the hypothesis [9], and at least ambiguous in
13 constructing one tree of life [10].

14 Though inconclusive, the body of evidence for the common descent of all life seems overwhelming.
15 However, we wish to point out that the notion of a single common ancestor is conceptually confusing
16 since it places some one individual organism on a special pedestal. It is as if LUCA were the first one
17 to depart in some significant way from its ancestors to be crowned as the earliest living being. To make
18 such a stark contrast between animate and inanimate should not be regarded merely as a logical lapse
19 that could be patched up by extending the notion of LUCA to an ancient community of organisms [11]
20 or even to a primordial class of biomolecules, but the search for a last common universal ancestor
21 indicates impaired understanding of what life actually is and how it came about [12,13].

22 **2. Evolution as a natural process**

23 To call living beings thermodynamic systems is nothing new [14], yet to claim that evolution is
24 nothing but a thermodynamic process seems to cock the eyebrows of many. The modern evolutionary
25 synthesis, by emphasizing the genetic mechanisms of evolution, narrows thinking from what evolution
26 actually is and distracts attention from comprehending what in fact powers evolution and what is the
27 profound basis of natural selection.

28 We argue here for the thermodynamic thinking of evolution by reminding that according to physics
29 everything that exists is associated with energy. Thereby nothing, irrespective of molecular, genetic or
30 any other mechanism, can escape the thermodynamic imperatives that force all systems toward
31 stationary-state in their environment. Hence evolution is regarded as any other process being driven by
32 a motive force, which in general terms is a difference in energy of any kind. Thus, a difference
33 between the energy of insolation and the energy in chemical potentials of molecular species [15] will
34 drive chemical reactions on Earth. Energy flows from heights to lows when substrates transform to
35 products. The flows of energy themselves will naturally select to channel via mechanisms, not only
36 molecular mechanisms but any kind of machinery, that will consume driving forces in least time [16].
37 The selection criterion, when given in terms of physics, is not ambiguous but the least-time
38 consumption of free energy is the unequivocal attribute of the fittest [17]. In the same way as a river
39 bed will run dry when water breaks a more effective channel, so will a suboptimal system lose its
40 supplies for a more effective system of species. The least-time consumption of free energy is the

1 natural bias of evolution. It is not a deterministic but holistic imperative because evolution, as it goes,
2 will alter its surroundings and thereby affect its own course. Hence forces of phylogenesis are physical.

3 It is worth to emphasize that the physical criterion of natural selection does not determine the course
4 of evolution because evolution itself will change the surrounding conditions which in turn will require
5 incessant re-evaluation of the least-time path. In other words, it is not the heterogeneity and complexity
6 of a system as such that would make its evolution intractable, but the non-computable character of
7 natural processes manifests itself already in the case of three bodies [18]. When there are two or more
8 alternatives to consume free energy, the natural process becomes intractable since the process itself
9 will affect its driving forces. Undoubtedly the heterogeneity and complexity contribute to the
10 computational challenge by providing various repositories of bound energy as well as numerous
11 combinations of pathways and mechanisms to consume differences in energy. As a result the complex
12 system may display at same time for a given sub-process both parallel and anti-parallel flows of
13 energy. In other words, in nature where everything depends on everything else, some supplies pile up
14 while others are consumed. The co-gradient and contra-gradient processes are customarily modeled,
15 e.g., by Lotka-Volterra equations, but accurately contained in the equation of evolution [19]. The net
16 absorption or emission of energy from the system to its surroundings displays the quest to attain
17 stationary state in the respective surroundings. The overall net dissipation stems from the universal
18 process that aims at attaining a stationary state with respect to “the zero-density surroundings”, i.e., the
19 heat death. The same thermodynamic imperative, however, will force diverse thermodynamic
20 potentials to pile up and adjust to each other, when surroundings are rich in energy such as here on
21 Earth that enjoys insolation.

22 This old, general formulation of evolution [20] is not only valid for biological processes. When
23 given by Newton’s 2nd law $\mathbf{F} = d\mathbf{p}/dt$, it is recognized to hold for any other natural process. Evolution,
24 when understood as a thermodynamic process, is not a random (stochastic) process at any level of
25 natural hierarchy, but it will keep on directing its momentum $\mathbf{p} = m\mathbf{v}$ along the resultant force to
26 diminish energy differences as soon as possible. The change in momentum $d\mathbf{p}/dt = m\mathbf{a} + \mathbf{v}dm/dt$
27 involves not only acceleration $m\mathbf{a}$ that balances the force within the system, but also the surroundings
28 in terms of dissipation of energy dE/dt in the form of a change in mass $dm/dt = dE/c^2dt$ to the free
29 space defined by the squared speed of light c^2 [17]. In other words, although an evolving system may
30 be subject to severe tension due to numerous non-parallel and even anti-parallel forces within the
31 system as well as with respect to the surroundings, it will processes along the resultant force. Also the
32 system-bound energy, i.e., heat capacity, is a supply that will resist imposed changes. It will manifest
33 itself as memory of the past trajectory. For example, the genome of an organism houses instructions of
34 the processes that were at least in the past circumstances effective in consuming free energy.
35 Accordingly organisms tend to proceed along the previously established paths of free energy
36 consumption just as cultures tend to follow their traditions in hopes of making the best use of their
37 characteristic circumstances [20].

38 The evolutionary equation [16] has also been known for a long time in its open, integral form as the
39 principle of least action [21]. It says that the flows of energy of any kind will track down from heights
40 to lows along the paths of least time. Over eons these paths of kinetic energy [22] have diverged for
41 maximal dispersal of energy. The resulting diversity is commonly depicted as a branching
42 phylogenetic tree.

1 **3. Inherent intractability**

2 The present-day parlance about evolution using only biological lingo does not make use of the
3 generality and rigor that is available when evolution is described by general terms in its entirety and
4 every detail as a probable physical process [23] of least-time free energy consumption. The concept of
5 energy can be assigned to everything that exists. Consequently the thermodynamic formalism will
6 ensure the conservation of quanta irrespective of mechanistic details and specific circumstances.
7 Notably, when evolution is given by the mathematical equation of motion [16,17,19], it can be
8 analyzed to realize that the path-dependent process cannot be solved. The paths of evolution cannot be
9 predicted forwards or traced backwards with certainty since the flows of energy and the differences of
10 energy as their driving forces cannot be separated from each other [16,17,19]. Since a natural process
11 is itself changing the surrounding conditions due to dissipation, the boundary conditions keep
12 changing. This is the fundamental reason why evolution is intractable when there are alternative paths
13 to consume free energy. In plain language, this means that when a species emerges with a new
14 characteristic, whether acquired by a mutation or new experience, its altered behavior [24] will in turn
15 exert forces on other species. According to thermodynamics evolution at all levels of natural hierarchy
16 [25] is a process that is itself carving the energy landscape and thereby directing its own course
17 [26,27]. There is no algorithm that could circumvent this innate character of non-computability [28].

18 **4. Ambiguous alignments**

19 Evolution is also intractable at the genomic level since information is physical too [29]. Ambiguity
20 about origin arises at the points of speciation where, mathematically speaking, derivatives of
21 evolutionary trajectories are inexact due to quantized dissipation [30]. In other words, information as
22 free energy [29] is invariably lost during evolution. Hence the past paths can no longer be traced with
23 certainty.

24 The non-deterministic nature of phylogenesis profoundly troubles attempts to align for ancestors.
25 To derive a cladogram even from a set of complete genome sequences of contemporary species is
26 inherently a non-computable task [28]. The non-computability manifests itself when searching for the
27 optimal alignment of one particular segment, because a positioning will affect the already adopted
28 alignments of other segments. That is, the computational problem keeps changing when it is being
29 solved [30], which is familiar from many other hard problems [32] such as protein folding [33].
30 Therefore, the calculated lineages, ranked in terms of probabilities, depend on the particular set of
31 aligned sequences, and no alignment can be announced as the actual. The search for the optimal
32 alignment is the search for the least-time path [34], but when the geodesic keeps changing by the
33 search itself, it obviously cannot be determined unambiguously.

34 Aligning for ancestry is in computational terms an NP-hard problem [35], which means that
35 increasing the number of sequences to be aligned causes an exponential growth in the computation
36 time needed to solve the problem. Circular reasoning is also a characteristic of sequence alignments;
37 sequences are aligned to build a tree, but to begin with the alignment itself needs a tree. The challenge
38 is to value diverse alignments and obtained trees [36] without having a firm deterministic basis. Even
39 the most effective algorithms [37] when modeling evolution, e.g., as if it were a Markovian process

[38,39], cannot remove the principle problem of intractable past [28]. Despite its non-computable character, evolution is not a random, stochastic process, but a physical process that directs along paths where free energy will be consumed in the least time. Emergence of biota itself is impressive evidence that circumstances, often referred to as boundary conditions, have changed when a huge amount of free energy from insolation has been bound to the system. Mathematically speaking, neither probabilities are invariant nor is there an ultimate norm in sight. Therefore, deterministic algorithms and unitary transformations are powerless to reconstruct phylogenesis precisely.

Moreover, when evolution is expressed in general thermodynamic terms, its mathematical analysis reveals, why populations of species from genes to galaxies display skewed, nearly lognormal distributions [40,41] as well as why evolution proceeds by punctuations and stasis [42] and why power laws are ubiquitous characteristics of both animate and inanimate processes and structures [18,43,44]. Since the least-time consumption of free energy as the universal criterion of natural selection renders evolution as a comprehensive and comprehensible process, the reference [45] “Nothing in biology makes sense except in the light of evolution.” is in place.

5. Conclusions

When evolution is understood as a thermodynamic process of least-time free energy consumption, it follows that phylogenetic analyses, as done using genetic information, will not be able to determine common ancestors. Moreover, when looking for LUCA, there is an inadvertently in-built impression that animate and inanimate would somehow be distinct from each other. This obscures comprehension of the principle of abiogenesis. Also the idea of aligning for a common ancestor is unintentionally a reductionist approach, but contemporary characters cannot be reduced one-to-one to primordial qualities because energy is not a constant of evolution. The modern evolutionary synthesis, by its gene-centered paradigm, narrows the view away from the general physical principle of evolution and thereby prevents obtaining a holistic understanding of nature as a unity.

References and Notes

1. Vanechoutte, M.; Fani, R. From the primordial soup to the latest universal common ancestor. *Research in Microbiology* **2009**, *160*, 437-440.
2. Hinegardner, R.T.; Engelberg, J. Rationale for a universal genetic code. *Science* **1963**, *142*, 1083-1085.
3. Schwartz, R.M.; Dayhoff, M.O. Origins of Prokaryotes, Eukaryotes, Mitochondria, and Chloroplasts. A perspective is derived from protein and nucleic acid sequence data. *Science* **1978**, *199*, 395-403.
4. Brown, J.R. Ancient horizontal gene transfer. *Nat. Rev. Genet.* **2003**, *4*, 121-132.
5. Christin, P.-A.; Weinreich, D.M.; Besnard, G. Causes and evolutionary significance of genetic convergence. *Trends in Genetics* **2010**, *26*, 400-405.
6. Doolittle, W.F. The practice of classification and the theory of evolution, and what the demise of Charles Darwin’s tree of life hypothesis means for both of them. *Phil. Trans. R. Soc. B* **2009**, *364*, 2221-2228.

- 1 7. Raup, D.M.; Valentine, J.W. Multiple origins of life. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 2981-
2 2984.
- 3 8. Theobald, D.L. A formal test of the theory of universal common ancestry. *Nature* **2010**, *465*, 219-
4 222.
- 5 9. Koonin, E.V.; Wolf, Y.I. The common ancestry of life. *Biol. Direct* **2010**, *18*, 64-68.
- 6 10. Doolittle, W.F.; Baptiste, E. Pattern pluralism and the Tree of Life hypothesis. *Proc. Natl. Acad.*
7 *Sci.* **2007**, *104*, 2043-2049.
- 8 11. Woese, C. The universal ancestor. *Proc. Natl. Acad. Sci.* **1998**, *95*, 6854-6859.
- 9 12. Lazcano, A. What is life? A brief historical overview. *Chem. Biodivers.* **2008**, *5*, 1-15.
- 10 13. Annala, A.; Annala, E. Why did life emerge? *Int. J. Astrobio.* **2008**, *7*, 293-300.
- 11 14. Monod, J. *Chance and Necessity: An Essay on the Nature of Philosophy of Modern Biology*. trans.
12 Wainhouse, A.; Collins: London, 1972.
- 13 15. Gibbs, J.W. *The scientific papers of J. Willard Gibbs*, Ox Bow Press: Woodbridge, CT, 1993-
14 1994.
- 15 16. Sharma, V.; Annala, A. Natural process – Natural selection. *Biophys. Chem.* **2007**, *127*, 123-128.
- 16 17. Kaila, V.R.I.; Annala, A. Natural selection for least action. *Proc. R. Soc. A* **2008**, *464*, 3055-3070.
- 17 18. Poincaré, J.H. Sur le problème des trois corps et les équations de la dynamique. Divergence des
18 séries de M. Lindstedt, *Acta Mathematica* **1890**, *13*, 1-270.
- 19 19. Mäkelä, T.; Annala, A. Natural patterns of energy dispersal. *Phys. Life. Rev.* **2010**, *7*, 477-498.
- 20 20. Annala, A.; Salthe, S. Cultural naturalism. *Entropy* **2010**, *12*, 1325-1343.
- 21 21. De Maupertuis, P.L.M. Accord de différentes loix de la nature qui avoient jusqu'ici paru
22 incompatibles. *Mém. Ac. Soc. Paris* **1744**, 417-426.
- 23 22. Du Châtelet, E. *Institutions de Physique*. Prault: Paris, 1740.
- 24 23. Whitfield, J. Survival of the Likeliest? *PLoS Biol.* **2007**, *5*, e142.
- 25 24. Anttila, J.; Annala, A. Natural Games. *Phys. Lett. A* **2011**, *375*, 3755-3761.
- 26 25. Salthe, S.N. *Evolving Hierarchical Systems: Their Structure and Representation*, Columbia
27 University Press: New York, 1985.
- 28 26. Van Valen, L. A new evolutionary law. *Evolutionary Theory* **1973**, *1*, 1-30.
- 29 27. Tuisku, P.; Pernu, T.K.; Annala, A. In the light of time. *Proc. R. Soc. A* **2009**, *465*, 1173-1198.
- 30 28. Annala, A. Physical portrayal of computational complexity. *ISRN Computational Mathematics*
31 **2011** (in press, <http://www.isrn.com/journals/cm/aip/321372/>) arXiv:0906.1084.
- 32 29. Karnani, M.; Pääkkönen, K.; Annala, A. The physical character of information. *Proc. R. Soc. A*
33 **2009**, *465*, 2155-2175.
- 34 30. Annala, A.; Salthe, S. Physical foundations of evolutionary theory. *J. Non-Equil. Thermodyn.*
35 **2010**, *35*, 301-321.
- 36 31. Milosavljevic, A. Discovering Dependencies via Algorithmic Mutual Information: A Case Study
37 in DNA Sequence Comparisons. *Machine Learning* **1995**, *21*, 35-50.
- 38 32. Applegate, D.L.; Bixby, R.E.; Chvátal, V.; Cook, W.J. *The Traveling Salesman Problem: A*
39 *Computational Study*, Princeton University Press: Princeton, NJ, 2006.
- 40 33. Sharma, V.; Kaila, V.R.I.; Annala, A. Protein folding as an evolutionary process. *Physica A* **2009**,
41 *388*, 851-862.

- 1 34. Owen, M.; Provan, J.S. A fast algorithm for computing geodesic distances in Tree Space.
2 *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2011**, *8*, 2-13.
- 3 35. Elias, I. Settling the intractability of multiple alignment. *J. Comput. Biol.* **2006**, *13*, 1323-1339.
- 4 36. Löytynoja, A.; Goldman, N. Uniting Alignments and Trees. *Science* **2009**, *324*, 1528-1529.
- 5 37. Huson, D.H.; Nettles, S.H.; Warnow, T.J. Disk-covering, a fast-converging method for
6 phylogenetic tree reconstruction. *J. Comput. Biol.* **1999**, *6*, 3-4.
- 7 38. Erdős, P.L.; Steel, M.A.; Székely, L.A.; Warnow, T.A. A few logs suffice to build (almost) all
8 trees (part 1). *Random. Struct. Algor.* **1999**, *14*, 153-184.
- 9 39. Daskalakis, C.; Roch, S. Alignment-Free Phylogenetic Reconstruction in Research. *in*
10 *Computational Molecular Biology, Lecture Notes in Computer Science* **2010**, *6044*, 123-137.
- 11 40. Limpert, E.; Stahel, W.A.; Abbt, M. Log-normal distributions across the sciences: keys and clues.
12 *Bioscience* **2001**, *51*, 341-352.
- 13 41. Grönholm, T.; Annala, A. Natural distribution. *Math. Biosci.* **2007**, *210*, 659-667.
- 14 42. Eldredge, N.; Gould, S.J. in *Models in Paleobiology*, ed. Schopf, T.J.M.; Freeman, Cooper: San
15 Francisco, CA, 1972; p. 82-115.
- 16 43. Bejan, A.; Marden, J.H. The constructal unification of biological and geophysical design. *Phys.*
17 *Life. Rev.* **2009**, *6*, 85-102.
- 18 44. Turcotte, D.L.; Pelletier, J.D.; Newman, W.I. Networks with side branching in biology. *J. Theor.*
19 *Biol.* **1998**, *193*, 577-592.
- 20 45. Dobzhansky, T. Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.*
21 **1973**, *35*, 125-129.

22
23 © 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article
24 distributed under the terms and conditions of the Creative Commons Attribution license
25 (<http://creativecommons.org/licenses/by/3.0/>).

26