Title

What is life and why, how and when did it begin?

Authors

Pierre M. Durand[1] and Richard E. Michod[2]
[1]University of the Witwatersrand, Johannesburg, South Africa and [2]University of Arizona, Tucson, USA

Abstract

The origin of life remains a compelling and poorly understood area in biology. Although various scenarios and explanations have been proposed, the fundamental questions such as what is life and why, how and when did it begin remain largely unanswered. The manuscript '**The First Gene** *Did Life Begin Following the Big Bang?*' by Joseph and Wickramasinghe offers an interesting genomic perspective, particularly for understanding when life emerged. Here, we deliver a commentary on this paper illustrating the complexities of origin of life research and indicate how genomics may assist in unravelling the myriad unanswered questions this field presents.

Introduction

The origin and evolution of life remains one of the most mysterious and challenging questions in biology. The field is fraught with stumbling blocks to such an extent that even a satisfactory definition for life and its origin that is acceptable to everyone seems unlikely (Luisi, 1998; Penny, 2005). The disagreements likely stem from the fact that there are researchers from diverse disciplines working in the field. This is undoubtedly a positive aspect, but what differentiates the living from non-living varies depending on what one considers to be the most essential feature of living systems. For biochemists life may be more about the origin of biochemical pathways, biomolecules like amino acids and nucleotides, or energy gradients; cell biologists are more concerned with the formation of proto-cell membranes; while evolutionary biologists focus on populations of the very first self-reproducing molecules to understand the origin of life. The explosion in whole genome data has allowed geneticists to use this information and ask questions about the "first" gene(s). This is the focus of the paper in question here "**The First Gene** *Did Life Begin Following the Big Bang?*" by Joseph and Wickramasinghe (2011).

Any avenue of enquiry that sheds light on the timing and circumstances around which life evolved is welcomed. At the outset; however, a familiar problem emerges. Joseph and Wickramasinghe ask the question "Did life begin following the big bang?" and attempt to answer it by estimating the timeframe for the origin of the first gene. The implicit assumption is that life originated with the evolution of the first gene. From a genetics perspective this may be an over-arching interest for understanding life's origins. Others, perhaps most, would disagree and have differing views on what constitutes life. As Penny (2005) has pointed out, before one can hope to uncover the fundamental questions concerning life's origins it is important to be explicit, even if only to list a number of features that, for a particular question, distinguish life from non-life.

Disentangling the origin of life questions

In addition to the problem of defining life, there are inter-related questions that are easily confused. On this point Joseph and Wickramasinghe are clearer. Their primary objective is to use genomic data to examine *when* the first gene arose. The question of *how* genes and genomes emerged, evolved and were duplicated is briefly considered at various points leading up to the statistical analyses (Figure 3, Joseph and Wickramasinghe, 2011) where the regression line describing gene and genome duplications is extrapolated to obtain an estimate for the origin of the first gene. The question of *why* life (or in this case the authors are really referring to first gene) arose is not considered. Focussing on a specific question is, of course, not a problem. However, the various issues are all connected and information related to one question informs and helps frame others. Ignoring the information available from other origin of life research questions can be problematic, which turns out to be the case here.

Our commentary relates information pertaining to the *why* and, to a lesser extent, *how* life evolved questions to Joseph and Wickramasinghe's work. However, there are two other issues that chemists and genomicists, respectively, may be particularly interested in. First, the study suggests that the first gene may have arisen soon after the Big Bang. Such a gene being perpetuated by protein-mediated replication would require significant elemental complexity. The authors specifically refer to genes (as opposed to RNA-mediated catalysis) and the study uses genomic information that relies on protein-mediated replication. The basic elements for nucleotides and simple amino acids (hydrogen, carbon, nitrogen, oxygen, phosphate etc.) would therefore have had to be present prior to the first gene emerging. How soon the basic elements existed following the Big Bang is unknown and it will be interesting to see whether the chemistry of the early universe before the formation of planets supports the timing for the origin of the first gene. Second, the study uses genetic information from extant genomes that cover <2 logarithmic units and the authors extrapolate the regression line a further 3 logarithmic units. This seems rather ambitious; but perhaps the major question for genomicists is that the authors assume a molecular clock-type of gradualism in gene and genome evolution. Whether this is a safe assumption is not clear. However, even over much smaller time scales when clock-like evolution is more likely, neutral evolution can be misleading when generating accurate dates and fossil calibration is essential (van Tuinen and Hadly, 2004).

Cooperation and conflict at the origin of life

The origin of the first gene is intimately associated with our understanding of the first replicating biomolecules. Despite many unanswered questions, Gilbert's RNA world hypothesis (Gilbert, 1986) and the groundbreaking work of Eigen and Schuster (Eigen and Schuster, 1977) on cooperation and conflict between early replicators enjoy considerable support for investigating problems at the beginning of the biotic world. These works are supported by empirical evidence describing the formation of ribonucleotide polymers by physio-chemical means (Engelhart and Hud, 2010) and are crucial for our understanding of the genetic basis for life. They describe a molecular living system long before a gene as described by Joseph and Wickramasinghe (evolved genetic code, transcription, translation etc) could have existed.

A comparative genomics perspective dating the origin of the first gene provides additional valuable information but ignores significant obstacles leading up to the formation of a gene; obstacles that would have taken time to overcome. Perhaps even more importantly, a protein-encoding first gene as inferred by the comparative genomics could not have existed at the beginning of life for several reasons. Before genes could have formed, there were two major problems that had to be overcome. These are the issues of informational error catastrophe and protein evolution. Of course amino acids and peptides may have emerged and co-evolved with catalytically active ribonucleotide polymers (ribozymes). However, as Eigen and others demonstrated, before any gene that functioned as part of a network or primitive genome could have existed, the issues of mutation rates and cooperation between replicators prevented the existence of genomes. This problem known as Eigen’s paradox: no genomes without enzymes and no enzymes without genomes; has received considerable attention. Solutions to the mutational meltdown of early replicators include Eigen and Schuster’s own hypercycle theory, which allows for shorter, catalytically inferior ribozymes to become functionally or replicatively connected through various-membered hypercycles. In addition, the concept of quasispecies (Eigen and Schuster, 1977, 1978 and 1979) and the importance of group theory and population structures (Michod, 1983); both of which are supported empirically (Arenas and Lehman, 2010), are important for understanding life’s origin.

The main point here is that our knowledge concerning the cooperative networks of ribonucleotide polymers at the origin of life means it was extremely unlikely for a gene as described by Joseph and Wickramasinghe to have emerged at the time of the Big Bang. Time was required for the evolution of groups of replicators, the genetic code and proteins before a well adapted gene or minimal genome existed. A further point worth considering is that, for the reasons of mutational error pointed out above, the original gene (protein coding or ribozyme-like) would have been short, perhaps <75 nucleotides (Eigen and Schuster, 1977). Comparing extant genomes encoding much larger, sophisticated protein replicative machinery to infer the origin of the inefficient replicators struggling for existence at the very beginning may well yield erroneous conclusions.

The potential role for genomics in origin of life research

As Joseph and Wickramasinghe show, despite the limitations inherent in genomic data, the field provides researchers with another avenue for exploring the origins of genes and genomes. Comparative genomics studies are central to understanding the emergence of protein-coding minimal genomes, which is particularly relevant for answering questions about the origin of cellular life. Another opportunity afforded by genomic data is the insight provided into the processes implicated by the emergence of ribozyme and gene networks. Far from being static, genomes are highly dynamic. Individual genetic elements demonstrate a range of evolutionary relevant behaviours and interactions (altruism, cooperation, competition etc.) (Durand and Michod, 2010; Foster, 2011). Understanding the life histories of selfish or cooperative mobile genetic elements in genomes can serve as a model system for understanding the socio-biology of early replicators. Genomic data can potentially be used in a multilevel selection framework (for example Takeuchi and Hogeweg, 2009) to understand the origin of the first genome and subsequent evolutionary transitions leading ever more complex life (Okasha, 2009).

Origin of life research – an interdisciplinary endeavor

Investigating the origin of life is a major challenge in biology. The field is still largely fragmented and divided along discipline specific lines to such a point that even a definition for life has been elusive. Before the mystery of life's early beginnings can be solved a cooperative effort between diverse fields is required, from mathematics and chemistry to evolutionary and molecular biology. Genomics provides a welcome new addition to this interdisciplinary endeavor.

Literature cited

Arenas, C. D., Lehman, N. (2010). Quasispecies like behaviour observed in catalytic RNA populations evolving in a test tube. BMC Evolutionary Biology 10: 80.
Durand, P. M., Michod, R. E. (2010). Genomics in the light of evolutionary transitions. Evolution 64: 1533-40.
Eigen, M., Schuster, P. (1977). The hypercycle: a principle of natural self-organization (Part A): emergence of the hypercycle. Naturwissenschaften 64: 541–565.
Eigen, M., Schuster, P. (1978). The Abstract hypercycle (Part B). Naturwissenschaften 65: 7-41.
Eigen, M., Schuster, P. (1979). The hypercycle: a principle of natural self-organization. Springer-Verlag, Berlin.
Engelhart, A. E. and Hud, N. V. (2010). Primitive genetic polymers. Cold Spring Harb Perspect Biol. 2: a002196.
Foster, K. R. (2011). The sociobiology of molecular systems. Nature Reviews Genetics 12: 193-203.
Gilbert, W. (1986). The RNA world. Nature 319: 618.
Joseph, R.J. and Wickramasinghe, N. C. (2011). The First Gene Did Life Begin Following the Big Bang? Journ Cosmol 16.
Luisi, P. L. (1998). About various definitions of life. Origins of life and evolution of biosphere 28: 613-622.
Michod, R.E. (1983). Evolution of the First Replicators. Amer. Zool. 23: 5-14.
Okasha, S. (2009). Evolution and levels of selection. Oxford University Press.
Penny, D. (2005). An interpretive review of the origin of life research. Biology and Philosophy 20: 633–71.

Takeuchi, N., Hogeweg, P. (2009). Multilevel Selection in Models of Prebiotic Evolution II: A Direct Comparison of Compartmentalization and Spatial Self-Organization. PLoS Comp Biol 5: 1-17.

Van Tuinen, M. and Hadly, E. A. (2004). Error in Estimation of Rate and Time Inferred from the Early Amniote Fossil Record and Avian Molecular Clocks. J Mol Evol 59:267–276.