ECRONICON OPEN ACCESS

EC MICROBIOLOGY
Research Article

Independent Genic-Encoded Enzymatic Reactions May Randomly Link into Multi-Step Biochemical Pathways in the Absence of Large Cell Selective Pressure

Waylen Teo[1,2], Zhao Jie Kwan[1,2], Aaron KY Lum[1,2], Shaunnessy MH Ng[1,2] and Maurice HT Ling[1,2,3]*

[1]*School of Life Sciences, Management Development Institute of Singapore, Singapore*

[2]*Department of Applied Sciences, Northumbria University, United Kingdom*

[3]*HOHY PTE LTD, Singapore*

***Corresponding Author:** Maurice HT Ling, HOHY PTE LTD, Singapore.

Received: January 03, 2024; **Published:** January 31, 2024

Abstract

Origin of metabolic pathways is an important milestone in evolution, linking pre-biotic/abiotic era to biotic era, and genic-encoded enzymatic reactions is at the core. Studies imply that *de novo* genic origin of early enzymes is plausible. There are also suggestions that large cells may be at the origin of life. However, the questions of whether unlinked enzymatic reactions link to multi-step biochemical pathways and whether large cells are necessary remain. Here, we use digital organisms to examine the emergence of multi-step biochemical pathways from independently genic-encoded enzymatic reactions. Our simulation results suggest that independently genic-encoded enzymatic reactions can randomly link into multi-step biochemical pathways in the absence of large cell selective pressure (p-value > 0.05). This suggests that genic-encoded multi-step biochemical pathways may arise randomly once enzymes are prevalent.

Keywords: *Genic-Encoded Enzymatic Reactions; Biochemical Pathways; Pre-Biotic/Abiotic Era; Biotic Era*

Introduction

How did primordial cells construct metabolic routes? - a question posed by Renato Fani [1], lies at the core of origin and evolution of metabolic pathways [2]. Several notable hypotheses have been put forth, as reviewed by Scossa and Fernie [3]. At the end of 20th century, Lazcano and Miller [4] proposed a semi-enzymatic approach to pathway origin which is still prominent after more than two decades [5]. The main feature of semi-enzymatic approach is that reactions may be catalyzed by a combination of enzymatic processes and non-biological/non-enzymatic processes. Non-enzymatic aspect is supported [6-8] by the discovery of non-enzymatic glycolysis and pentose phosphate pathway [9,10] and more recently, the three-step reversal oxaloacetate into succinate can be achieved without enzymes and using hydrogen as reducing agent in biological conditions [11]. On the enzymatic front, primitive enzymes may arise independently or in combination as deoxyribozymes, ribozymes, or peptide enzymes [12] through abiotic polymerizations [13] as short random chains of peptides may contain putative protein domains [14] and long DNA sequences may contain putative genes [15,16]. This is further supported by studies suggesting that promoters [17,18] and coding sequences [19] may originate from random DNA sequences.

Since *de novo* origin of functional genes is possible [15-19], the genic origin of early enzymes is plausible. In addition, cell size may also be important as hyperthermophilic giant cells may lie at the origin of life [20], which is supported by the recent discovery of bacterium that is nearly 1 centimetre long [21]. Hence, by relaxing the criteria on the origin of metabolic pathways from semi-enzymatic [4] to only consider enzymatic approach, the questions of whether unlinked enzymatic reactions can link to multi-step biochemical pathways and whether large cells are necessary remain. However, these questions are not feasible to be examined experimentally. Digital organisms (DOs), which are computer-simulated organisms [22-24] and had been used to explore various experimentally challenging evolutionary scenarios [25-34] with alternative metabolisms [35]. DOs has been considered as instances of life rather than simulations of life [36] and by extension, can be considered alive [37]. In this study, we use DOs to study the possibility of independently genic-encoded enzymatic reactions in large cells to link into multi-step biochemical pathways. Our simulation results suggest that independently genic-encoded enzymatic reactions can randomly link into multi-step biochemical pathways without the need for large cells.

Materials and Methods

Simulation system: Digital Organism Simulation Environment (DOSE) [38-40] with D2-plus genomic interpreter was used as the simulation platform. D2-plus was modified from D2 Interpreter [33]. The modifications were as follow: Firstly, a null gene was defined as two zero nucleotides. Secondly, genes 71 to 99 were defined as undefined genes, which resembles pseudogenes. Lastly, mutation is at the nucleotide level (where two nucleotides encode for a gene) rather than gene level.

Simulation setup: Each simulation replicate had 100 DOs, simulated for 1000 generations. The genome of each DO was initiated with 2000 bases of zero, which was equivalent to 1000 null genes. Mutation rate was set at 1% or equivalent to 20 base mutations per DO per generation. The selection process from one generation to the next is based on previous studies [16-18,32,33]. Briefly, the lowest decile of the organisms by fitness were removed per generation but in event where more than 50% of the population could be removed, a random selection of 10 organisms were removed instead. After removal, a random selection of remaining DOs after removal were replicated to top up the population to 100 DOs for the next generation. Four fitness criteria, where three of them acting as large cell selective pressure, were defined: (a) In metabolite sum trial, fitness was defined as the sum of intracellular metabolites. (b) In metabolite average trial, fitness was defined as the amount of average metabolites across all 25 metabolites. (c) In unique metabolite trial, fitness was defined as the number of distinct metabolites. (d) In control trial, no fitness criterion was defined, and 50 DOs were randomly removed in the selection process. Each DO is given an empty cytoplasm (all intracellular metabolites were set to zero) before executing the metabolism coded by its genome. All 25 environmental metabolites were set to 1000.

Analysis: Analysis of simulation results was performed using three metrics defined in DOSSIER [41]: (a) MeasureSum function was used to count the total number of enzymatic genes within the genome of a DO, which refers to the amount of genome coding for enzymes. (b) MeasureDiversity function was used to count the unique number of enzymatic genes within the genome of a DO, which refers to the number of enzymes within the DO. (c) MeasureEfficiency function was used to measure the global efficiency [42] of the DO's metabolic network, which was a proxy measure to the presence and efficiency of information flow within the metabolic pathways. Each metric for each DO per generation per replicate was averaged as mean metric per generation per replicate, which was then used to calculate the grand mean and standard error of metric per generation.

Results and Discussion

Semi-enzymatic chemistry [4,5] in large cells [20] may lie at the origin of metabolic pathways. By relaxing the approach and only consider enzymatic origin of pathways, we examines the possibility where independently unlinked enzymatic reactions link to form biochemical pathways and whether large cells are required. Three large cell selective pressure were defined; namely, the sum of intracellular metabolites (metabolite sum), the amount of average metabolites (metabolite average), and the number of distinct metabolites

(metabolite unique). Three metrics were defined for each DO; namely, number of enzymatic genes, number of unique enzymatic genes, and metabolic efficiency.

Our simulation results show three important results. Firstly, the grand mean of enzymatic genes (Figure 1A), grand mean of gene diversity (Figure 1B), and grand mean of metabolic efficiency (Figure 3) are not significantly different (p-value > 0.05) in all 3 large cell selective pressures. This suggests that the effects of large cell selection have no significant impact (p-value > 0.05) on the eventual genome and metabolism. Secondly, the selective pressure for the number of unique metabolites found in each DO appears not to result in large cells as compared to selective pressures for either high quantity of total metabolites or high average metabolite quantity (Figure 2A). Given the same osmolarity, this selection results in smaller cells compared to control (p-value < 0.001) by half other simulated generations. Nevertheless, selecting for the number of unique metabolites does result in the expected outcome of higher number of distinct metabolites in DOs as compared to other selective pressures (Figure 2B, p-value < 0.001). Thirdly and most importantly, all 3 metrices are not significantly different (p-value > 0.05) between selective pressures and the random, no-selection control; after around 450 generations.

Our simulation results also show increased number of distinct metabolites in the DOs at the end of 1000 generations (Figure 2B). From 700th generation, control selection plateaus at about 14.6 types of metabolites per DO while metabolite sum and metabolite average trials plateaus at about 15.4 and 15.5 types of metabolites per DO, respectively. The mean number of distinct metabolites in metabolite sum and metabolite average trials are not significantly different from each other (p-value > 0.05) but significantly higher than control (p-value < 0.001). Selection for unique metabolites resulted in the highest metabolite variety, at the mean of 18.9 per DO after 700th generation, which is significantly higher than large cell trials and control (p-value < 0.001).

Crucially, our simulation results show that the grand mean of global efficiency rapidly increase to more than 0.4 within the first 50 generations and plateaus at about 0.41 after 200 generations for all selections (Figure 3). The grand mean of global efficiency at 200th generation is significantly higher (p-value < 0.001) than that at 50th generation. The degree of metabolite connectively plateaus at 41% of that of a fully connected metabolic network based on the definition of global efficiency [42,43]. In this context, a fully connected metabolic network is a metabolism where any 2 metabolites can be connected by a single reaction. This suggests that even at 58.4% (14.6 out of 25 types of metabolites) of the possible metabolite usage in control trial, linking of independent reactions into reaction networks is plausible. Hence, this suggests that independently genic-encoded enzymatic reactions has the potential to randomly link into multi-step biochemical pathways, in the absence of large cell selection pressure and not significantly different (p-value > 0.05) with or without large cell selection pressure. Therefore, genic-encoded multi-step biochemical pathways has the potential randomly once enzymes are prevalent. Yet, this does not imply that large cells have no impact in evolutionary processes and may still be relevant in the origin of eukaryotes [44,45].

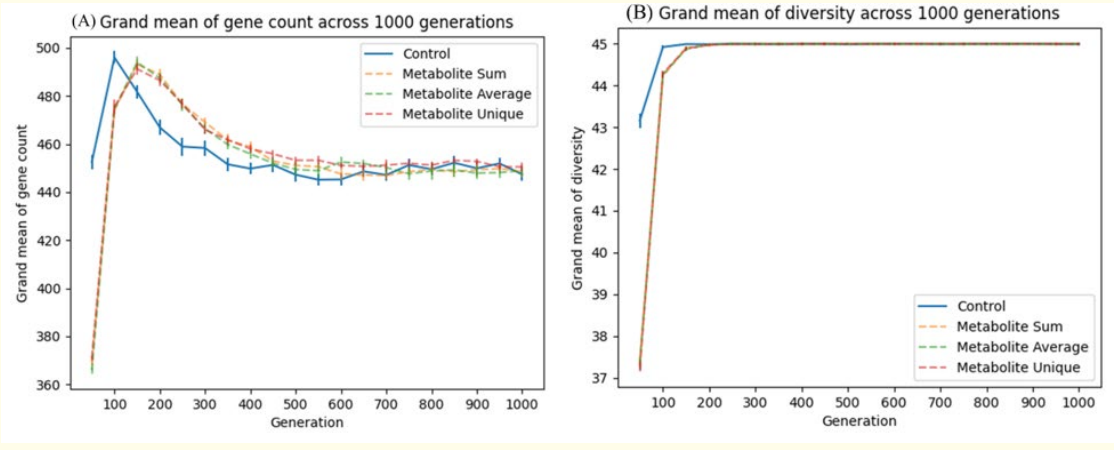

Figure 1: *Grand mean of the number of enzymatic genes and gene diversity in DOs over generations. Panel A shows the grand mean of the number of enzymatic genes. Panel B shows the grand mean of gene diversity.*

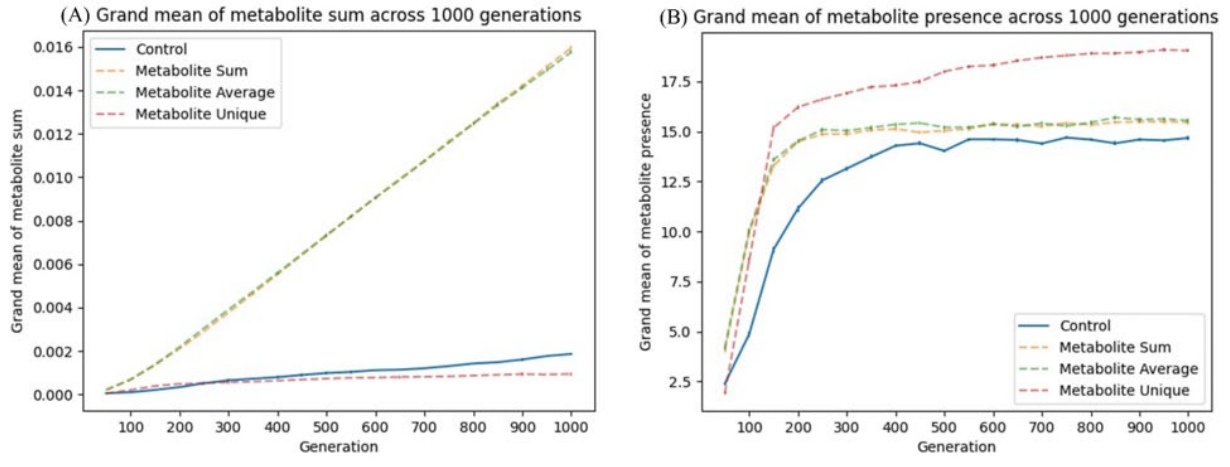

Figure 2: *Grand mean of the quantity of metabolites and metabolite usage in DOs over generations. Panel A shows the grand mean of the quantity of metabolites. Panel B shows the grand mean of the types of metabolites found (representing metabolite usage).*

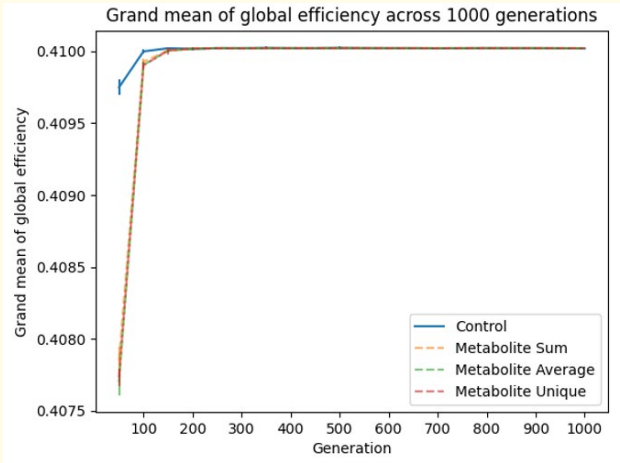

Figure 3: *Grand mean of metabolic efficiency in DOs over generations.*

Conclusion

Using DOs, we show that genic-encoded multi-step biochemical pathways may arise randomly, in the absence of large cell selection, once enzymes are prevalent.

Supplementary Materials

Codes for this study can be downloaded at https://bit.ly/PathwayEmergence.

Conflict of Interest

The authors declare no conflict of interest.

Bibliography

1. Fani R. "The origin and evolution of metabolic pathways: why and how did primordial cells construct metabolic routes?" *Evolution: Education and Outreach* 5.3 (2012): 367-381.
2. Fani R and Fondi M. "Origin and evolution of metabolic pathways". *Physics of Life Reviews* 6.1 (2009): 23-52.
3. Scossa F and Fernie AR. "The evolution of metabolism: how to test evolutionary hypotheses at the genomic level". *Computational and Structural Biotechnology Journal* 18 (2020): 482-500.
4. Lazcano A and Miller SL. "On the origin of metabolic pathways". *Journal of Molecular Evolution* 49.4 (1999): 424-431.
5. Becerra A. "The semi-enzymatic origin of metabolic pathways: inferring a very early stage of the evolution of life". *Journal of Molecular Evolution* 89.3 (2021): 183-188.
6. Keller MA., *et al*. "The widespread role of non-enzymatic reactions in cellular metabolism". *Current Opinion in Biotechnology* 34 (2015): 153-161.
7. Muchowska KB., *et al*. "Nonenzymatic metabolic reactions and life's origins". *Chemical Reviews* 120.15 (2020): 7708-7744.
8. Prosdocimi F., *et al*. "Prebiotic chemical refugia: multifaceted scenario for the formation of biomolecules in primitive earth". *Theory in Biosciences* 141.4 (2022): 339-347.
9. Keller MA., *et al*. "Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible archean ocean". *Molecular Systems Biology* 10.4 (2014): 725.
10. Ralser M. "An appeal to magic? the discovery of a non-enzymatic metabolism and its role in the origins of life". *The Biochemical Journal* 475.16 (2018): 2577-2592.
11. Rauscher SA and Moran J. "Hydrogen drives part of the reverse krebs cycle under metal or meteorite catalysis". *Angewandte Chemie International Edition* 61.51 (2022): e202212932.
12. Tagami S and Li P. "The origin of life: RNA and Protein co-evolution on the ancient earth". *Development, Growth and Differentiation* 65.3 (2023): 167-174.
13. Sim BK and Ling MH. "Possibility of abiotic genesis of biochemistry". *EC Microbiology* 16.6 (2020): 104-109.
14. Ling MH. "*De novo* putative protein domains from random peptides". *Acta Scientific Microbiology* 2.4 (2019): 109-112.
15. Thong-Ek C., *et al*. "Potential *de novo* origins of archaebacterial glycerol-1-phosphate dehydrogenase (G1PDH)". *Acta Scientific Microbiology* 2.6 (2019): 106-110.
16. Kwek BZ., *et al*. "Random sequences may have putative beta-lactamase properties". *Acta Scientific Medical Sciences* 3.7 (2019): 113-117.
17. Ardhanari-Shanmugam KD., *et al*. "*De novo* origination of *Bacillus subtilis* 168 promoters from random sequences". *Acta Scientific Microbiology* 2.11 (2019): 07-10.
18. Usman S., *et al*. "*Pseudomonas balearica* DSM 6083T promoters can potentially originate from random sequences". *MOJ Proteomics and Bioinformatics* 8.2 (2019): 66-70.

19. Neo CY and Ling MH. "Prevalence and length of open reading frames vary across randomly generated sequences of different nucleotide compositions". *EC Microbiology* 16.7 (2020): 72-78.

20. Stetter KO. "Hyperthermophiles in the history of life". *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* 361.1474 (2006): 1837-1842; discussion 1842-1843.

21. Volland J-M., *et al*. "A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles". *Science* 376.6600 (2022): 1453-1458.

22. Langton CG. "Studying artificial life with cellular automata". *Physica D: Nonlinear Phenomena* 22.1-3 (1986): 120-149.

23. Elena SF and Sanjuán R. "The effect of genetic robustness on evolvability in digital organisms". *BMC Evolutionary Biology* 8 (2008): 284.

24. Ortega R., *et al*. "Ontology for the avida digital evolution platform". *Scientific Data* 10.1 (2023): 608.

25. Anderson CJR and Harmon L. "Ecological and mutation-order speciation in digital organisms". *The American Naturalist* 183.2 (2014): 257-268.

26. Castillo CFG and Ling MHT. "Resistant traits in digital organisms do not revert preselection status despite extended deselection: implications to microbial antibiotics resistance". *BioMed Research International* (2014): 648389.

27. Ling MH. "Applications of artificial life and digital organisms in the study of genetic evolution". *Advances in Computer Science: An International Journal* 3.4 (2014): 107-112.

28. Yao Y., *et al*. "Using digital organisms to study the evolutionary consequences of whole genome duplication and polyploidy". *PloS One* 14.7 (2019): e0220257.

29. Castillo CF., *et al*. "Resistance maintained in digital organisms despite guanine/cytosine-based fitness cost and extended de-selection: implications to microbial antibiotics resistance". *MOJ Proteomics and Bioinformatics* 2.2 (2015): 00039.

30. Wilke CO and Adami C. "The biology of digital organisms". *Trends in Ecology and Evolution* 17.11 (2002): 528-532.

31. Chew SS., *et al*. "Rapid genetic diversity with variability between replicated digital organism simulations and its implications on cambrian explosion". *EC Clinical and Medical Case Reports* 3.11 (2020): 64-68.

32. Ang DG and Ling MH. "Sudden and steep harsh environment results in over-compensation in digital organisms". *EC Microbiology* 17.7 (2021): 104-113.

33. Sooriya Kannan KS., *et al*. "Nutrient availability impacts intracellular metabolic profiles in digital organisms". *Acta Scientific Microbiology* 6.5 (2022): 18-25.

34. Dolson E and Ofria C. "Digital evolution for ecology research: a review". *Frontiers in Ecology and Evolution* 9 (2021): 750779.

35. Mozhayskiy V and Tagkopoulos I. "Microbial evolution *in vivo* and *in silico*: methods and applications". *Integrative Biology* 5.2 (2013): 262-277.

36. O'Neill B. "Digital evolution". *PLoS Biology* 1.1 (2003): E18.

37. Koh YZ and Ling MH. "On the liveliness of artificial life". *iConcept Journal of Human-Level Intelligence* 3 (2013): 1.

38. Castillo CF and Ling MH. "Digital organism simulation environment (DOSE): A library for ecologically-based in silico experimental evolution". *Advances in Computer Science: An International Journal* 3.1 (2014): 44-50.

39. Castillo CF and Ling MH. "Digital organism simulation environment (DOSE) Version 1.0.4". Current STEM, Volume 1 (Nova Science Publishers, Inc.) (2018): 1-106.

40. Ling MH. "Ragaraja 1.0: The genome interpreter of digital organism simulation environment (DOSE)". The Python Papers Source Codes 4 (2012): 2.

41. Maitra A and Ling MH. "DOSSIER: A toolkit to extract data from digital life simulations using dose". *Acta Scientific Computer Sciences* 4.7 (2022): 37-40.

42. Latora V and Marchiori M. "Efficient behavior of small-world networks". *Physical Review Letters* 87.19 (2001): 198701.

43. Latora V and Marchiori M. "Economic small-world behavior in weighted networks". *The European Physical Journal B - Condensed Matter and Complex Systems* 32.2 (2003): 249-263.

44. Schavemaker PE and Muñoz-Gómez SA. "The role of mitochondrial energetics in the origin and diversification of eukaryotes". *Nature Ecology and Evolution* 6.9 (2022): 1307-1317.

45. Williams RJP. "A system's view of the evolution of life". *Journal of the Royal Society, Interface* 4.17 (2007): 1049-1070.

Volume 20 Issue 2 February 2024