

Regime Shift of Genome Size Crossing the Chasm of Eukaryogenesis

Hai-Jun Wang^{1†}, Xu Zhao^{1†}, Chi Xu², Yang Liu³, Ying Liu¹, Lei Shi¹, Erik Jeppesen^{4, 5, 6, 7, 8}, Ping Xie^{9, 1*}

¹Institute for Ecological Research and Pollution Control of Plateau Lakes, School of Ecology and Environmental Science, Yunnan University; Kunming, China

²School of Life Sciences, Nanjing University; Nanjing, China

³Shenzhen Key Laboratory of Marine Microbiome Engineering, Institute for Advanced Study, Shenzhen University; Shenzhen, China

⁴Department of Bioscience, Aarhus University; Silkeborg, Denmark

⁵Sino-Danish Centre for Education and Research; Beijing, China

⁶Limnology Laboratory, Department of Biological Sciences, Middle East Technical University; Ankara, Turkey

⁷Centre for Ecosystem Research and Implementation (EKOSAM), Middle East Technical University; Ankara, Turkey

⁸Institute of Marine Sciences, Middle East Technical University; Mersin, Turkey

⁹Donghu Experimental Station of Lake Ecosystems, State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences; Wuhan, China

†These authors contributed equally to this work.

* Corresponding author.

Email: wanghaijun@ynu.edu.cn; zhaoxu@mail.ynu.edu.cn; xuchi@nju.edu.cn; yangliu@szu.edu.cn; liuy35@hotmail.com; shilei1225@foxmail.com; ej@bios.au.dk; xieping@ihb.ac.cn

Abstract: The origin of the nucleus remains a great mystery in life science, although nearly two centuries have passed since the discovery of nuclei. To date, studies of eukaryogenesis have focused largely on micro-evolutionary explanations. Here, we examined macro-patterns of C-values (the total amount of DNA within the haploid chromosome set of an organism) for over 110,000 species and the chromosome numbers for over 11,000 species and their potential links with the state of atmospheric oxidation over geological time. Eukaryogenesis was in sync with an over 2.5 order-of-magnitude increase in genome size from prokaryote to eukaryote, and also with a rapid rise of atmospheric oxidation, suggesting that eukaryogenesis would have resulted from a regime shift of genomes driven by the oxidation-driven complexification and structuralization (e.g. chromatin packing).

Keywords: eukaryogenesis; genome complexification; atmospheric oxidation; macroevolution

1. Introduction

Eukaryogenesis—the entire process by which the defining traits of eukaryotic cells arose in the lineage that eventually gave rise to all present-day eukaryotes—has long been of great interest to evolutionary biologists ^{1,2}, with multiple hypotheses proposed to date (e.g., various symbiotic models and the autogenous and exomembrane hypotheses) (table S1) ^{3,4}. Undoubtedly, eukaryogenesis should be a product of both macro- and micro-evolutions adapting to the changing environments. Although almost certainly related to origination of the nucleus in some way, it is not clear how to evaluate the relation between evolutionary complexification (increase in size) of genomes and eukaryogenesis. The acquisition of mitochondria capable of producing adenosine triphosphate (ATP) through oxidative phosphorylation has been proposed as prerequisite to eukaryogenesis ⁵⁻⁷. However, little is known about the possible link between genome complexification and atmospheric oxidation (with high energy efficiency), and its potential role in eukaryogenesis. Since this study focused mainly on genomic evolution from prokaryotes to eukaryotes, a relatively simple typology of five biological kingdoms was applied, although more sophisticated typologies do exist.

In the present study, we have collected C-values for 110,431 species (covering bacteria, archaea, fungi, plants, and animals) (text S1) as well as chromosome number data for 11,792 species and explored their potential evolutionary relationships with the state of atmospheric oxidation over geological time. The term “C-value” refers to the amount (in picograms) of DNA contained within a haploid nucleus of an organism, being widely used to reflect the expected genetic complexity of a species, and in practice the term is often used interchangeably with genome size ⁸. A chromosome is a complex linear organization of a DNA molecule aided by various packaging proteins (histones) and chaperone proteins. Each chromosome contains part of or all the genetic material of an organism. In a tiny cell, genomic complexification and structuralization were likely inseparable processes, which perhaps led to the emergence of eukaryotes upon reaching a threshold (Fig. 1).

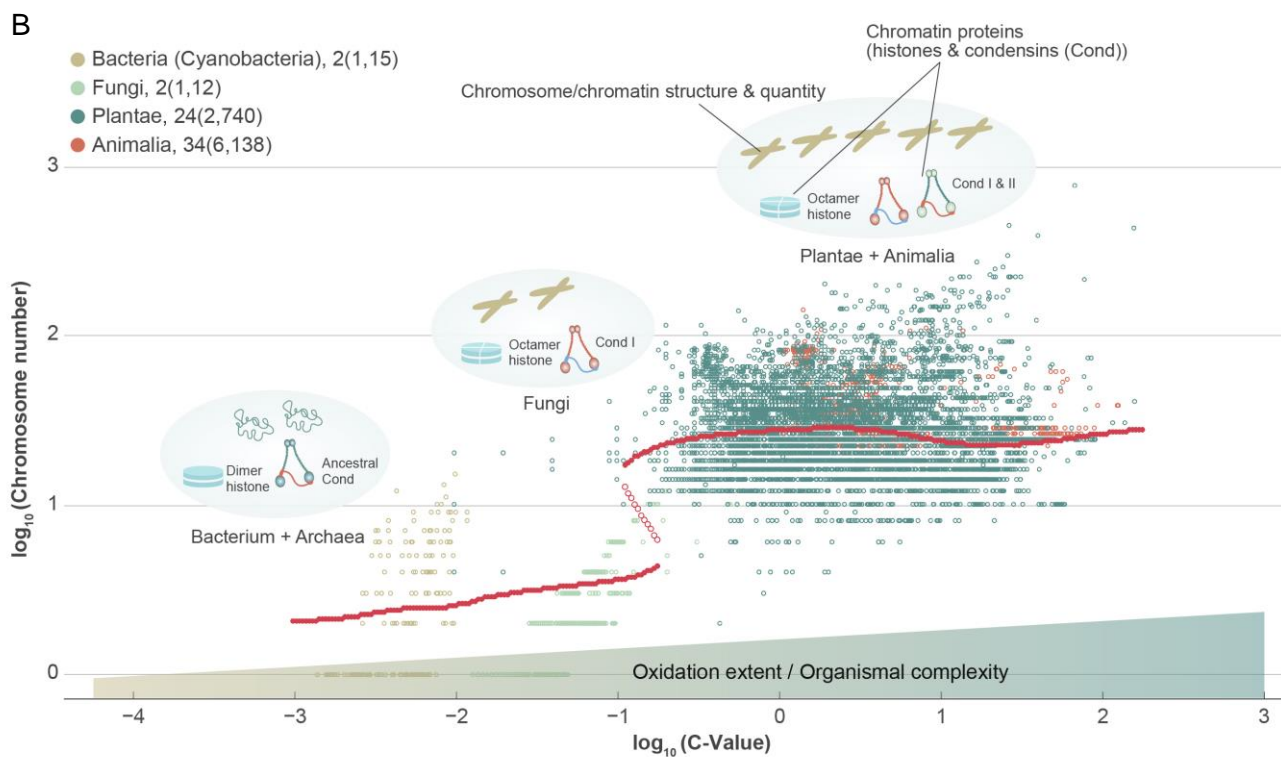
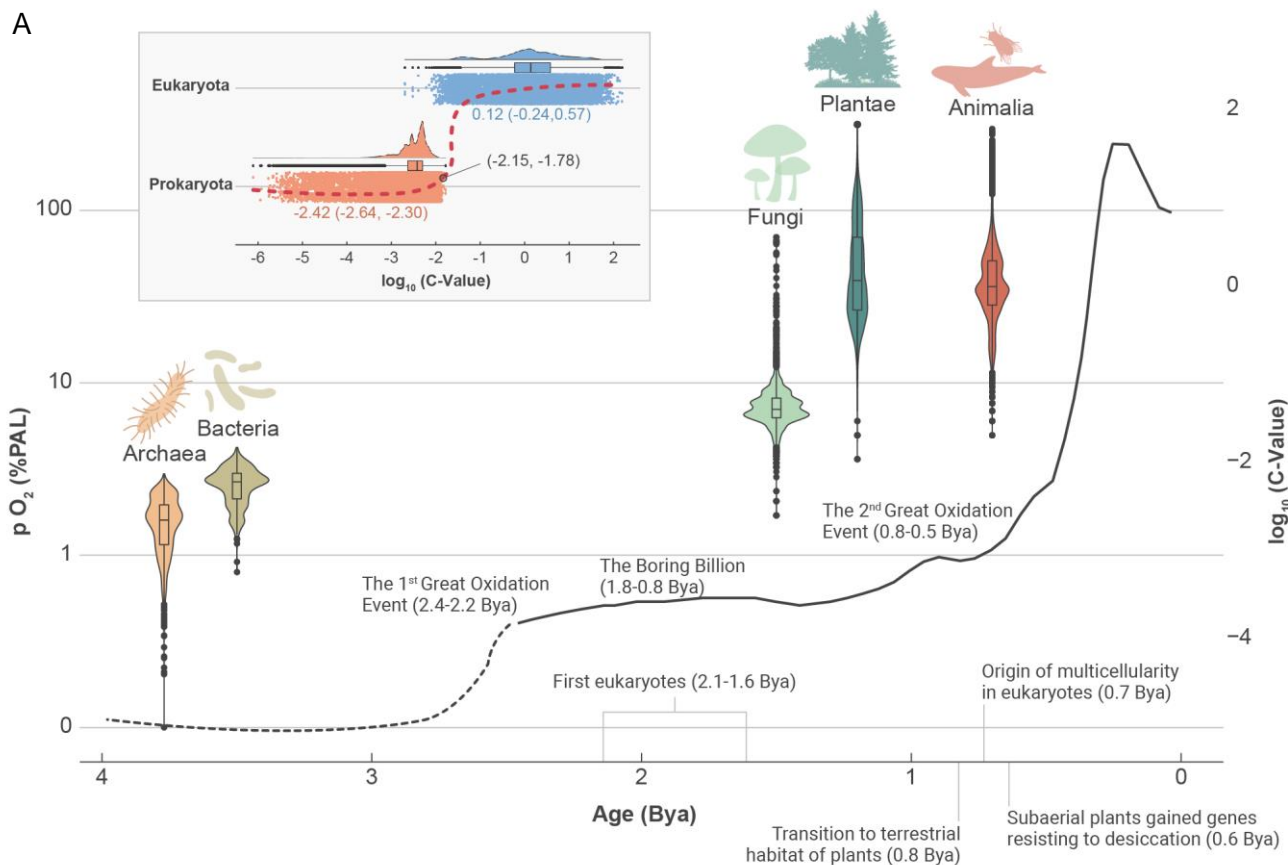


Fig. 1. Macro patterns of genome complexity (C-values) for the evolutionary lineages of various biological kingdoms. (A) C-value distributions of the kingdoms and percentage atmospheric oxygen concentration (pO²) along with geological time. The first appearance time was estimated at ca. 3.77 Bya for archaea ¹⁶, 3.5 Bya for cyanobacteria ¹⁷, 1.5 Bya for fungi ¹⁸, 1.2 Bya for plants ¹⁹, and 0.7 Bya for animals ²⁰. The C-value distributions of prokaryotes and eukaryotes show their medians (1st quartile, 3rd quartile), with blue open circle and value indicating the thresholds (95%-100%) for the transition from prokaryotes to eukaryotes. **(B) Relationship of C-value with chromosome number with fitting lines based on a potential landscape model.** The data after the kingdoms shows the corresponding median (minimum, maximum) of the chromosome numbers. It should be noted that chromosomes for cyanobacteria are a kind of unstable circular form between chromatin and real chromosome ²¹. The evolution of structure, quantity, and packing-associated proteins of chromosomes were also given along a gradient of organismal complexity and oxidation extent.

2. Materials and Methods

The approach of Potential Landscape Model (also termed as stability landscape) ⁹ has been used to infer the existence of stable states across a given state space from noisy data^{10, 11}. This approach assumes an underlying stochastic system with a potential function:

$$dz = -U'(z)dt + \sigma dW$$

where z represents the focal state variable (here chromosome size), $U(z)$ represents the potential function, σ represents the noise level, and dW is a noise term. The equilibrium values were calculated by determining the local minima and maxima of the resulting potential landscape. In Fig 1B, the red curve (local minima of the potential landscape) represents the inferred stable equilibrium, and the open circles (local maxima of the potential landscape) represent unstable equilibrium.

Web sources for data of C-values and chromosome numbers are as follows. Bacteria and archaea ¹²: <https://img.jgi.doe.gov/> ; Fungi ¹³: <http://www.zbi.ee/fungal-genomesize/>; Plants ¹⁴: <https://cvalues.science.kew.org/> ; Animals ¹⁵: <http://www.genomesize.com>.

3. Results and discussion

There has been a consensus that atmospheric oxidation had profound impacts on the evolution of life on Earth. Changes in the atmospheric oxygen level (pO_2) and the C-value distributions of the five kingdoms over geological time are presented in Fig. 1A²²⁻²⁵. The median C-value shows large differences between prokaryotes and eukaryotes, with the lowest being for archaea ($\log_{10}(C) = -2.76$) and bacteria (-2.41) (LC group), the intermediate for fungi (-1.38) (MC group), and the highest for animals (0.16) and plants (0.23) (HC group) (Fig. 1A, table S2). According to the estimated first appearance time of each kingdom^{16-18, 20, 26, 27}, the LC group corresponds to the low pO_2 period, the MC group to the moderate pO_2 period, long after the First Great Oxidation Event (2.4 billion years ago, Bya), and the HC group to a period slightly before or during the Second Great Oxidation Event (0.8-0.5 Bya). The origination of life on Earth (3.85 Bya) and the appearance of the first eukaryote (2.1-1.6 Bya) shows at least a lag of 1.75 Bya²⁸. However, the lag is only 0.3 Bya from fungus to plants^{18, 19}, and 0.5 Bya from plants to animals^{19, 20}, suggesting the potential role of atmospheric oxidation in promoting diversification and genome complexity, and thus speciation. Our view is supported by the idea that energy production by the acquisition of mitochondria could have enabled eukaryogenesis^{5, 6, 29}.

Interestingly, the C-value distribution of prokaryotes (including bacteria and archaea) shows evidence of heavy crowding at the high-value end, while scattered at the low-value end and at both ends of the three eukaryotic kingdoms (Fig. 1A). This suggests an apparent barrier of upper C-values for prokaryotes, i.e., only if a C-value threshold was passed could a shift from prokaryote to eukaryote take place. If taking the highest 5% value of this high-value end of prokaryotes, the C-value threshold for the prokaryote-to-eukaryote shift can be defined as $-2.15 \sim -1.78$ ($\log_{10}x$).

The relationship between chromosome number and C-values showed a sigmoid shape (Fig. 1B). The test by a potential landscape model according to Livina et al. (2010)⁹ suggested two contrasting states, with fungi and bacteria (represented by cyanobacteria, the taxa with chromosome number reported) in the low-value group, while plants and animals in the high-value group. This suggests that the transition from the lower to the higher group takes the shape of a regime shift. Additionally, the fact that eukaryotes have much larger genome sizes than prokaryotes suggests that accidental symbioses could not make a nucleus. In other words, the

genomic complexity of eukaryotes could be greatly increased only if the evolution of the nucleus had experienced a long sequential process.

Related evolution of the structure (i.e., circular vs. linear), quantity (numbers), and packing-associated proteins of chromosomes along with the organismal complexity and oxidation extent was further summarized (Fig. 1B). The atmospheric oxidation likely promoted complexification and structuralization of genomes and hence triggered the origination of the nucleus after a sharp increase of genomic size upon crossing a threshold, as supported by the following facts or inferences. First, atmospheric oxidation provided the energetic basis for genomic complexification^{30,31}. Second, there exists an upper limit of genomic size for prokaryotes. Third, formation of nuclear membranes and increase in chromosome number enabled orderly management and control of the complex biochemical system with increasing genomic complexity. Fourth, some proteins neofunctionalized to help highly efficient chromatin packing^{32,33}. Associated packing proteins are mainly simple nucleoid-associated ones in prokaryotes, whereas various complex ones (e.g., histones, cohesins, etc.) are present in eukaryotes^{33,34}. Eukaryotic chromosomes are therefore highly condensed and linked in an orderly manner to ensure that the genome fits inside the nucleus. For example, the chromosome packing ratio is as high as 10,000:1 for human beings when comparing the length of uncoiled DNA to that of a chromosome in metaphase (Annunziato, 2008); this ratio can be as high as 60,400 for some plant species³⁵ and over 65,600 for fungi³⁶ based on the length of chromosome and its genome contents³⁵.

4. Conclusions

In conclusion, a regime shift of genome size occurred crossing the chasm of eukaryogenesis, and genomic complexification and structuralization (especially chromatin packing) driven by elevated atmospheric oxidation was likely a key macro-evolutionary force for the long-lasting eukaryogenesis that had been collaboratively completed through natural selection by microevolution (e.g., symbiosis and random mutation). Eukaryogenesis paves the way for rapid speciation, leading to adaptive radiation of various life forms on Earth. Further efforts should be made to integrate multidisciplinary evidence to reveal in more detail how processes of different evolutionary levels may have interacted synergistically to complete eukaryogenesis with sequential stepwise hierarchies.

References

1. López-García, P. & Moreira, D. Selective forces for the origin of the eukaryotic nucleus. *BioEssays* **28**, 525–533 (2006).
2. Dacks, J. B. et al. The changing view of eukaryogenesis – fossils, cells, lineages and how they all come together. *J. Cell Sci.* **129**, 3695–3703 (2016).
3. de Roos, A. D. G. The Origin of the Eukaryotic Cell Based on Conservation of Existing Interfaces. *Artif. Life* **12**, 513–523 (2006).
4. Baum, D. A. A comparison of autogenous theories for the origin of eukaryotic cells. *Am. J. Bot.* **102**, 1954–1965 (2015).
5. Lane, N. & Martin, W. The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
6. Lane, N. Energetics and genetics across the prokaryote-eukaryote divide. *Biol. Direct* **6**, 35 (2011).
7. Koonin, E. V. Energetics and population genetics at the root of eukaryotic cellular and genomic complexity. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 15777–15778 (2015).
8. Gregory, T. R. Macroevolution, hierarchy theory, and the C-value enigma. *Paleobiology* **30**, 179–202 (2004).
9. Livina, V. N., Kwasniok, F. & Lenton, T. M. Potential analysis reveals changing number of climate states during the last 60 kyr. *Clim. Past.* **6**, 77–82 (2010).
10. Hirota, M., Holmgren, M. & Van Nes, E. H. Scheffer, M. Global resilience of tropical forest and savanna to critical transitions. *Science* **334**, 232–235 (2011).
11. Xu, C. et al. Remotely sensed canopy height reveals three pantropical ecosystem states. *Ecology* **97**, 2518–2521 (2016).
12. Mukherjee, S. et al. Genomes Online database (GOLD) v.8: overview and updates. *Nucleic Acids Res.* **49**, D723–D733 (2020).
13. Kullman, B., Tamm, H. & Kullman, K. Fungal Genome Size Database. <http://www.zbi.ee/fungal-genomesize/> (2005).
14. Leitch, I. J., Johnston, E., Pellicer, J., Hidalgo, O. & Bennett, M. D. Plant DNA C-values Database (Release 7.1). <https://cvalues.science.kew.org/> (2019).
15. Gregory, T. R. Animal genome size database. <http://www.genomesize.com/> (2021).
16. Dodd, M. S. et al. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature* **543**, 60–64 (2017).
17. Schopf, J. W. Fossil evidence of Archaean life. *Philos. Trans. R. Soc. B-Biol. Sci.* **361**, 869–885 (2006).
18. Martin, W. et al. Early Cell Evolution, Eukaryotes, Anoxia, Sulfide, Oxygen, Fungi First (?), and a Tree of Genomes Revisited. *IUBMB Life* **55**, 193–204 (2003).
19. Retallack, G. J. & Lawson, J. D. Fossil soils as grounds for interpreting the advent of large

- plants and animals on land. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.* **309**, 105–142 (1985).
20. Valentine, J. W. The Evolution of Multicellular Plants and Animals. *Sci. Am.* **239**, 140–158 (1978).
 21. Jain, I. H., Vijayan, V. & O'Shea, E. K. Spatial ordering of chromosomes enhances the fidelity of chromosome partitioning in cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13638–13643 (2012).
 22. Schopf, J. W. Microfossils of the Early Archean Apex Chert: New Evidence of the Antiquity of Life. *Science* **260**, 640–646 (1993).
 23. Feng, D. F., Cho, G. & Doolittle, R. F. Determining divergence times with a protein clock: Update and reevaluation. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 13028–13033 (1997).
 24. Vellai, T. & Vida, G. The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* **266**, 1571–1577 (1999).
 25. Liu, Y. et al. Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* **593**, 553–557 (2021).
 26. Butterfield, N. J. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386–404 (2000).
 27. Cheng, S. et al. Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution. *Cell* **179**, 1057–1067.e14 (2019).
 28. Holland, H. D. Evidence for life on Earth more than 3850 million years ago. *Science* **275**, 38–39 (1997).
 29. Lane, N. *The Vital Question: Why is Life the Way it is?* (Profile Books, London, 2015)
 30. Planavsky, N. J. et al. A case for low atmospheric oxygen levels during Earth's middle history. *Emerg. Top. Life Sci.* **2**, 149–159 (2018).
 31. Xie, P. Who is the “Matchmaker” between Proteins and Nucleic Acids? *Innov.* **2**, 1–35 (2021).
 32. Luijsterburg, M. S., White, M. F., van Driel, R. & Dame, R. T. The Major Architects of Chromatin: Architectural Proteins in Bacteria, Archaea and Eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* **43**, 393–418 (2008).
 33. Dame, R. T., Rashid, F.-Z. M. & Grainger, D. C. Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.* **21**, 227–242 (2020).
 34. Hirano, T. Condensin-Based Chromosome Organization from Bacteria to Vertebrates. *Cell* **164**, 847–857 (2016).
 35. Guha, A., Sinha, R. K. & Sinha, S. Average packing ratio as a parameter for analyzing the karyotypes of dioecious cucurbits. *Caryologia* **57**, 117–120 (2004).
 36. Poma, A., Venora, G., Miranda, M. & Pacioni, G. The karyotypes of three *Tuber* species (Pezizales, Ascomycota). *Caryologia* **55**, 307–313 (2002).

Acknowledgements We thank Profs. Eugene V. Koonin (National Institutes of Health, USA), Meng Li (Institute for Advanced Study, Shenzhen University, China), Pei-Yun Cong (Institute of Palaeontology, Yunnan University, China) for their valuable comments and Dr. Anne Mette Poulsen (Aarhus University) for linguistic assistance. This research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31000000), the National Natural Science Foundation of China (32061143014), and the Yunnan Provincial Department of Science and Technology (202001BB050078). HJW was supported by the Youth Innovation Association of the Chinese Academy of Sciences as an excellent member (Y201859). EJ was supported by the Tübitak Outstanding Researchers Program, BIDEB 2232 (118C250).

Author contributions HJW and PX designed the research. HJW, XZ, LS, and YiL collected the data and performed the analyses of C-value distributions. CX and LS performed the potential landscape model analysis. HJW, XZ, and YiL prepared the first draft of the manuscript, and PX, CX, YaL and EJ contributed greatly to its improvement.

Competing interests The authors declare no competing interests.

Supplementary Information

Table S1 A list of the prevalent hypotheses on the origin of nucleus

Name	Connotation	Reference
Viral Eukaryogenesis Membrane Fusion Hypothesis	The eukaryotic nucleus evolved from a complex DNA virus. The viral ancestor of the nucleus is proposed to have been a complex enveloped DNA virus similar to the present-day poxviridae/ASF viruses, except that its host was an archaeon. It possessed a large linear chromosome and could enter the archaeal host cell by membrane fusion.	37
The Viral Eukaryogenesis (VE) Hypothesis	The first eukaryotic cell was a multimember consortium consisting of a viral ancestor of the nucleus, an archaeal ancestor of the eukaryotic cytoplasm, and a bacterial ancestor of the mitochondria, a cell wall-less archaeon and an α -proteobacterium established a syntrophic relationship, and then a complex DNA virus permanently lysogenized the archaeal syntrophy to produce a consortium of three organisms that evolved into the eukaryotic cell.	38
Updated Viral Eukaryogenesis (VE) Hypothesis	The eukaryotic nucleus evolved from the viral factory of a DNA virus that infected the archaeal ancestor of the eukaryotes.	39
Virus Origin Symbiotic Hypothesis	Large DNA viruses, such as poxviruses, led to the emergence of the eukaryotic cell nucleus. Posing symbiotic contact of an orthopoxvirus ancestor with an archaebacterium.	41

Updated Viral Origin Symbiotic Hypothesis	An ancestral virus of currently existing giant viruses of PAM (Phycodnaviridae, Mimiviridae-related viruses, Mimiviridae, Asfarviridae, Pandoraviruses/ Mollivirus) superclade constructed a VF (viral factory) in the cytoplasm of proto-eukaryotic cells, which was surrounded by a cytoplasmic inner membrane (putatively ER). The host cell developed a membrane compartmentalization to protect genomic DNA from viral attacks.	41
Endosymbiotic Models	The nucleus derives from the incorporation of an archaeon within a bacterium. A eubacterial host that engulfed an archaeobacterial endosymbiont that underwent a transformation into the nucleus (a) The key event in the origin of the eukaryotic cell is postulated to be the engulfment of an 'eocyte' archaeobacterium by a Gram-negative eubacterium that presumably lacked a cell wall. (b) As the membrane of the host surrounded the guest species, its own membrane (containing ether-linked lipids) became redundant and was lost. (c) Eventual separation of the internalized membrane from the plasma membrane led to the formation of the nuclear envelope and the endoplasmic reticulum. The formation of these new compartments was preceded or accompanied by duplication of the genes for the chaperone proteins (e.g. hsp70, hsp90, etc.). The transfer of the host genome to the newly formed nucleus, and an assortment of the genes from the two parents, led to the formation of the ancestral eukaryotic cell.	42
Chimeric Symbiotic Models	Archaea and bacteria would be the two only survivor lines descending from the last common ancestor and eukaryotes would derive from the merging of the two. The incorporation of the future mitochondrion to an archaeal host triggered eukaryogenesis.	1
Searcy's Hypothesis	Sulfur-mediated symbiosis between a wall-less, sulfur-respiring Thermoplasma-like archaeon and photo- or chemoautotrophic H ₂ S-utilizing bacterium.	43
Hydrogen Hypothesis	A hydrogen-mediated symbiosis between a hydrogenoclastic methanogenic archaeon and a hydrogen-producing alphaproteobacterium, with hydrogen being used to reduce the CO ₂ also released by the bacterium for methanogenesis.	43
HS Syntrophy Hypothesis	A complex sulfate-reducing deltaproteobacterium (host). An endosymbiotic hydrogen-producing Asgard-like archaeon (future nucleus). A metabolically versatile, facultatively aerobic, sulfide-oxidizing and potentially anoxygenic photosynthesizing alphaproteobacterium (future mitochondrion).	43
Entangle–Engulf–Endogenize (E3) Model	The host of anaerobic archaea is symbiotic with a kind of sulfate-reducing bacteria. Symbiotic bacteria are receptors of fatty acids and hydrogen produced by amino acid metabolism of the host. Subsequently, the host and another aerobic bacteria symbiosis to form mitochondria, involving three species of prokaryotes. A dual symbiosis, in microoxic environments, of an Asgard archaeon that degraded amino acids to short-chain fatty acids and hydrogen with a sulfate-reducing bacterium (SRB) and an aerobic organotrophic alphaproteobacterium that scavenged toxic O ₂ . As the consortium progresses towards increasingly oxic zones, the interaction with the alphaproteobacterium becomes stronger until it is engulfed. The SRB symbiosis is transient and eventually lost.	44

Reverse Flow Model: An Updated Symbiogenetic Model	A metabolic syntrophy between anaerobic ancestral Asgard archaea and facultative anaerobic alphaproteobacteria has provided the selective force for the establishment of a stable symbiotic interaction that has subsequently led to the origin of the eukaryotic cell.	45
Updated hydrogen hypothesis	Hydrogen could transfer from the bacterial symbiont δ -Proteobacterium, which is a fermentative organoheterotroph to a hydrogen-dependent autotrophic Asgard archaea (Wukongarchaeota) host.	12
Syntrophic Chimeric Fusion Hypothesis	A chimeric cell evolved via symbiogenesis by syntrophic merger between an archaeobacterium and a eubacterium. The archaeobacterium, a thermoacidophil, generated hydrogen sulfide to protect the eubacterium. By eubacterial-archaeobacterial genetic integration, the chimera, an amitochondriate heterotroph. This “earliest branching protist” that formed by permanent DNA recombination generated the nucleus as a component of the karyomastigont.	46
The Symbiosis of Pyrococcus into γ -Proteobacteria	The nucleus emerged from an archaeal endosymbiont (Pyrococcus -like), which was engulfed by a γ -proteobacterium.	47
Symbiogenic Theory	A symbiosis between a spirochaete and an archaeobacterium without a cell wall (most likely Thermoplasma-like in her view), leading to both the eukaryotic flagellum and the nucleus.	48
Koonin and Yutin's Hypothesis	the archaeal ancestor of eukaryotes was a complex form, rooted deeply within the TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota) superphylum, that already possessed some quintessential eukaryotic features, in particular, a cytoskeleton, and perhaps was capable of a primitive form of phagocytosis that would facilitate the engulfment of potential symbionts	49
Coevolutionary Theory	The origin of the nucleus depended on the prior evolution of a primitive endomembrane system and a primitive mitosis, both brought about by and associated with the origin of phagocytosis. The concerted origins of the endomembrane system and cytoskeleton, subsequently recruited to form the cell nucleus and coevolving mitotic apparatus.	50
Exocytosis Hypothesis	The first endomembranes to evolve during eukaryote evolution had secretory, and not phagocytic, function. Eukaryogenesis was initiated not by phagocytosis, but by the evolution of secretion by exocytosis, with phagocytosis arising later. The establishment of a secretory endomembrane system happened before or in parallel with the evolution of a nucleus.	51

Cavalier-Smith's Hypothesis	Both the nucleus and the endoplasmic reticulum come from invagination of the plasma membrane. Eukaryotic cells evolve directly from bacteria and do not need endosymbiosis. An autogenous (non-symbiotic) origin of a phagocytosing amitochondriate eukaryote (an archezoan) via point mutational changes leading to a host that does not need a mitochondrion at all to enjoy its phagocytotic lifestyle, but acquires one nonetheless.	52
Autogenous Models	Nucleus and cytoplasm evolved from a single prokaryotic lineage.	4
Gould and Dring's Hypothesis	Endospore formation of Gram-positive bacteria resulted in the origin of the nucleus. The protoplast of a single cell divides during endospore formation in such a manner that the cell engulfs a portion of its own cytoplasm, which then becomes surrounded by a double membrane resulting in the cell's nucleus.	53
Exomembrane Hypothesis	Nucleus originated from a single early cell which then evolved into a second outer cell membrane.	3

References

37. Bell, J. L. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? *J. Mol. Evol.* **53**, 251-256 (2001).
38. Bell, J. L. The viral eukaryogenesis hypothesis: a key role for viruses in the emergence of eukaryotes from a prokaryotic world environment. *Ann. N.Y. Acad. Sci.* **1178**, 91-105 (2009).
39. Bell, J. L. Evidence supporting a viral origin of the eukaryotic nucleus. *Virus Res.* **289**, 198168 (2020).
40. Takemura, M. Poxviruses and the Origin of the Eukaryotic Nucleus. *J. Mol. Evol.* **52**, 419-425 (2001).
41. Takemura, M. Medusavirus Ancestor in a Proto-Eukaryotic Cell: Updating the Hypothesis for the Viral Origin of the Nucleus. *Front. Microbiol.* **11**, 571831 (2020).
42. Gupta, S. & Golding, G. B. The origin of the eukaryotic cell. *Trends Biochem.Sci.* **21**, 166-171 (1996).
43. López-García, P. & Moreira, D. The Syntrophy hypothesis for the origin of eukaryotes revisited. *NAT. MICROBIOL* **5**, 655-667 (2020).
44. Imachi, H. et al. Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**, 519-525 (2020).
45. Spang, A. et al. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *NAT. MICROBIOL* **4**, 1138-1148 (2019).

46. Margulis, L., Dolan, M. F. & Guerrero, R. The chimeric eukaryote: Origin of the nucleus from the karyomastigont in amitochondriate protists. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6954-6959 (2000).
47. Horiike, T., Hamada, K., Miyata, D. & Shinozawa, T. The Origin of Eukaryotes Is Suggested as the Symbiosis of Pyrococcus into γ -Proteobacteria by Phylogenetic Tree Based on Gene Content. *J. Mol. Evol.* **59**, 606-619 (2004).
48. Margulis, L., Chapman, M., Guerrero, R. & Hall, J. The last eukaryotic common ancestor (LECA): acquisition of cytoskeletal motility from aerotolerant spirochetes in the Proterozoic Eon. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13080-13085 (2006).
49. Koonin, E. V. & Yutin, N. The Dispersed Archaeal Eukaryome and the Complex Archaeal Ancestor of Eukaryotes. *Cold Spring Harbor Perspect. Biol.* **6**, a016188 (2014).
50. Cavalier-Smith, T. Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. *Biol. Direct* **5**, 7 (2010).
51. Jékely, G. Small GTPases and the evolution of the eukaryotic cell. *BioEssays* **25**, 1129-1138 (2003).
52. Cavalier-Smith, T. The origin of eukaryotic and archaeobacterial cells. *Ann. N.Y. Acad. Sci.* **503**, 17-54 (1987).
53. Gould, G. W. & Dring, G. J. On a possible relationship between bacterial endospore formation and the origin of eukaryotic cells. *J. Theor. Biol.* **81**, 47-53 (1979).

Table S2. Basic characteristics of C-values for various biological kingdoms

	<i>n</i>	Minimum	1 st quartile	Median	Mean	3 rd quartile	Maximum
Prokaryotes	91287	7.86×10^{-7}	2.30×10^{-3}	3.80×10^{-3}	3.77×10^{-3}	5.01×10^{-3}	1.65×10^{-2}
Bacteria	89273	7.86×10^{-7}	2.36×10^{-3}	3.86×10^{-3}	3.81×10^{-3}	5.03×10^{-3}	1.65×10^{-2}
Archaea	2014	4.52×10^{-6}	8.62×10^{-4}	1.74×10^{-3}	1.93×10^{-3}	2.70×10^{-3}	6.60×10^{-3}
Eukaryotes	22424	2.00×10^{-3}	5.8×10^{-1}	1.32	4.34	3.68	1.52×10^2
Fungi	2404	2.00×10^{-3}	3.30×10^{-2}	4.20×10^{-2}	8.38×10^{-2}	5.80×10^{-2}	5.93
Plants	12017	1.00×10^{-2}	7.30×10^{-1}	1.70	5.40	5.90	1.52×10^2
Animals	8003	2.00×10^{-2}	8.40×10^{-1}	1.43	4.03	3.01	1.33×10^2