"III Congreso Latinoamericano de Astrobiología (2021)"
Revista Mexicana de Astronomía y Astrofísica Serie de Conferencias (RMxAC), **55**, 56–58 (2023)

https://doi.org/10.22201/ia.14052059p.2023.55.14

III Latin American Congress of Astrobiology (3CLA)(Virtual-Modality, August 3-6, 2021)
Editors: Leticia Carigi, Sandra I. Ramírez Jiménez, Miguel Chávez Dagostino, and Millarca Valenzuela - DOI: https://doi.org/10.22201/ia.14052059p.2023.55.14

THE GENOMES OF COMPLEX MULTICELLULAR ORGANISMS ON EARTH ARE CHARACTERIZED BY HIGH INTRON-RICHNESS

I. Lozada-Chávez[1], A. N. Lozada-Chávez[2], P. F. Stadler[1,3,4,5,6], and S. J. Prohaska[1]

RESUMEN

El origen y evolución de la multicelularidad compleja (MC) en la Tierra es de particular interés en la Astrobiología para la búsqueda de vida extraterrestre compleja en el Universo. Sin embargo, la MC terrestre ha evolucionado de forma independiente en muy pocos linajes eucariontes: algas cafés y rojas, plantas, animales y hongos. Paradójicamente, la complejidad de los eucariontes no se correlaciona con su tamaño de genoma ni con su número de genes, sino que ésta parece correlacionar con la expansión del ADN que no codifica para proteínas (ncADN), tal como los intrones y las moléculas regulatorias de ARN. Aquí presentamos una definición formal de la MC y algunos resultados que muestran que los genomas de los organismos MC son "ricos" en intrones, independientemente del tamaño de su genoma. Ésto apoya nuestra hipótesis de que la MC en la Tierra es una transición evolutiva mayor promovida, entre otros factores, por la presencia de innovaciones en la estructura del genoma y la diferenciación celular debido a la evolución convergente de ciertas formas de ncADN.

ABSTRACT

The origin and evolution of complex multicellularity (CM) on Earth is of particular interest for Astrobiology to search for complex alien life elsewhere in the Universe. However, CM is restricted in our planet to eukaryotes and has evolved independently a few times within red and brown algae, plants, animals, and fungi. Paradoxically, the complexity of eukaryotes does not correlate with genome size or the total number of genes, but it seems to correlate with the expansion of DNA that does not encode for proteins (ncDNA), such as spliceosomal introns and regulatory non-coding RNAs. Here, we present a formal definition for CM and summarize some findings showing that the genomes of CM organisms and their closest ancestral relatives are indeed characterized by high intron-richness, regardless their genome size. These findings support our hypothesize that CM on Earth is the outcome of major evolutionary transitions involving, among other factors, the presence of innovatory changes in genome structure and cell differentiation due to the convergent evolution of specific ncDNA classes.

Key Words: eukaryotes — exobiology — genome evolution — introns — multicellularity — non-coding DNA

1. INTRODUCTION

Several studies estimate that the probability to detect complex extraterrestrial life is much smaller in comparison with the detection of microbial-like life (Foucher et al. 2017). Complex macroscopic and intelligent life on Earth has originated *via* transitions to ***multicellularity***, which we define here as the phenotype characterized by the self-organization of cells that undergo a transition in individuality to perform cooperative consumption of energy, survival, and reproduction in some cases. Multicellularity has arisen multiple times during the evolution of the three domains of life (Grosberg & Strathmann 2007), and it can be induced in experimental settings (Ratcliff et al. 2013). However, multicellularity is hypothesized to unfold into two different evolutionary transitions: *simple* or *complex*. Complex multicellularity (CM) is a major evolutionary transition and its convergent origins are restricted to a few independent lineages in eukaryotes (Knoll 2011; Lozada-Chávez et al. 2011): laminarian brown algae (Phaeophyta), florideophyte red algae (Rhodophyta), eumetazoan animals (Cnidaria, Bilateria), land plants (Charophyta, Bryophyta, Tracheophyta), and fungi (Basidiomycota, Ascomycota).

Paradoxically, the organismal complexity in eukaryotes does not correlate with their 60,000-fold range of genome size or their total number of genes; instead, it seems to correlate with the expansion

[1] Evo-Devo & Bioinformatics Group, Department of Computer Science, University of Leipzig, Härtelstrasse 15-18, D–04107 Leipzig, Germany (ilozada@bioinf.uni-leipzig.de).

[2] Department of Biology and Biotechnology, University of Pavia, via Ferrata, 27100 Pavia, Italy.

[3] Max Planck Institute for Mathematics in the Sciences, Insel Str. 22, D-04103 Leipzig, Germany.

[4] Fraunhofer Institute for Cell Therapy and Immunology, Perlick Str. 1, D-04103 Leipzig, Germany.

[5] Department of Theoretical Chemistry, University of Vienna, Währinger Str. 17, 1090 Vienna, Austria.

[6] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM, 87501, USA.

III Latin American Congress of Astrobiology (3CLA)(Virtual-Modality, August 3-6, 2021)
Editors: Leticia Carigi, Sandra I. Ramírez Jiménez, Miguel Chávez Dagostino, and Millarca Valenzuela - DOI: https://doi.org/10.22201/ia.14052059p.2023.55.14

of DNA that do not encode for proteins (ncDNA) in the genome, such as spliceosomal introns, regulatory non-coding RNAs, repeats and pseudogenes (Lozada-Chávez et al. 2011). However, the evolutionary roles of ncDNA in Eukarya remain an enigma, specially intriguing are spliceosomal introns. Introns are the non-protein coding sequences of the genes that need to be removed through the process known as "splicing" to form a functional protein or RNA molecule. Introns are ubiquitous and sometimes highly abundant in the genomes of eukaryotes (see examples in Table 1), without providing an obvious functional or selective benefit to the organism (Lozada-Chávez et al. 2018). Such conundrum motivates us to ask: are introns merely a consequence of changes in genome size? Is there any relationship between intron-richness and CM evolution?

2. METHODS AND RESULTS

We use here phylogenetically based comparative analyses for 461 eukaryotes from all major supergroups to examine whether key features defining body plan development and genome complexity can distinguish between simple (SMOs) and complex multicellular organisms (CMOs).

2.1. *A definition of Complex Multicellularity*

Typically, the number of differentiated cell types (UCTs) is used as the defining feature of CM. However, accurate estimates of UCTs are only available for a small number of species, and they also fail to capture the underlying principles driving multicellular complexity (Niklas et al. 2014). Based on previous work (Rensing 2016; Nagy 2018), we created three "working definitions" to classify species in our dataset, first as unicellular or multicellular, and then as simple or complex multicellular. To that end, we compiled from literature four criteria embracing key aspects of cellular development and life cycle for the 461 eukaryotes: (i) presence or absence of multiple and differentiated cell types, (ii) aggregative or clonal development, (iii) facultative or obligate multicellularity, and (iv) development of reversible or irreversible tissue-based body plans. We thus define ***complex multicellularity*** as the phenotype (organism) exhibiting an irreversible transition in individuality produced by tissue-based body plans, through the developmental commitment of multiple and different cell types originated from a common cell-line ancestor. Based on this approach, we classified 77 species as unicellular, 96 species as SMOs, and 288 species as CMOs. Definitions for unicellularity and simple multicellularity are described in Lozada-Chávez et al. 2018.

2.2. *Genome size and intron-richness in complex multicellular organisms*

By running our program `GenomeContent` across 461 complete genomes of eukaryotes, we systematically calculated the distribution of their genome-wide features (number, size, density, genome-content, repetitive composition) for protein-coding sequences (exons) and three ncDNA classes: introns, unique ncDNA and repeats. We then calculated evolutionary correlations among all genome-wide features and both genome size and multicellular complexity at the broadest and local phylogenetic scales. We controlled for phylogenetic non-independence in our comparative analyses by using Principal Components Analysis (`phyl.pca`, R-package: `phytools`) and Phylogenetic Generalized Least Squares (`pgls`, R-package: `caper`). To account for the phylogenetic uncertainty of the (true but unknown) Tree of Eukaryotes, we performed our evolutionary correlations with alternative tree topologies based on supertree construction, NCBI–based taxonomy, and protein-domain content (`PFAM` database and `hmmer` program).

Surprisingly, we found that CMOs and their closest ancestral relatives are characterized by high intron-richness, regardless their genome size. Indeed, our results demonstrate that variations in the features measuring length and repeat composition of introns are only weakly correlated with changes in genome size at the broadest phylogenetic scale ($r^2 < 0.4$, $p < 0.001$), while features measuring intron abundance (within and across genes) do not scale at all. In remarkable contrast, genome size correlates strongly and positively with repetitive sequences ($r^2 > 0.7$, $p < 0.001$) but negatively with several protein-coding features. However, the strength of these correlations can fluctuate at the lineage-specific level. Our findings are robust to phylogenetic uncertainties, variation in species number, gene annotations and genome size estimations. Some results are roughly summarized in Table 1 & Figure 1, and fully described in Lozada-Chávez et al. 2018.

3. DISCUSSION AND CONCLUSIONS

Contrasting pioneering studies (Lynch et al. 2011), our results do not support a concerted evolution between genome size and ncDNA at large phylogenetic scales in eukaryotes. Instead, our findings unveil ncDNA —and particularly introns— as a dynamically evolving genetic trait of Eukarya, very probably under the influence of several life-history factors and evolutionary (adaptive and non-adaptive) forces. Our results endorse a genetic distinction between SM and CM in Archaeplastida and

III Latin American Congress of Astrobiology (3CLA)(Virtual-Modality, August 3-6, 2021)
Editors: Leticia Carigi, Sandra I. Ramírez Jiménez, Miguel Chávez Dagostino, and Millarca Valenzuela - DOI: https://doi.org/10.22201/ia.14052059p.2023.55.14

TABLE 1

ESTIMATED VALUES OF ORGANISMAL AND GENOMIC COMPLEXITY FOR SELECTED SPECIES OF THIS STUDY[a]

Species	Taxa	MT	UCT	GS	GN	ID	IS	IC
Dictyostelium discoideum	Amoebozoa	SM	1 – 6	34	13, 089	1.86	137.47	6.21
Ectocarpus siliculosus	Stramenophila	CM	4 – 14	214	16, 276	7.3	772.55	36.04
Chondrus crispus	Rhodophyta	CM	6 – 7	105	9, 603	1, 99	180.28	0.47
Volvox carteri	Chlorophyta	SM	1 – 2	138	15, 285	7.24	401.09	27.22
Physcomitrella patens	Bryophyta	CM	11 – 26	518	38, 354	5.18	302.05	6.05
Arabidopsis thaliana	Tracheophyta	CM	5 – 44	156	27, 206	5.49	187.72	11.29
Neurospora crassa	Ascomycota	SM	5 – 28	38.64	9, 820	2.07	134.07	5.69
Amanita muscaria	Basidiomycota	CM	9 – 30	40.7	18, 153	4.29	82.42	10.34
Nematostella vectensis	Cnidaria	CM	3 – 22	224.94	27, 273	6.11	1002.83	40.49
Drosophila melanogaster	Protostomia	CM	10 – 70	180	13, 924	3.94	706.53	20.98
Homo sapiens	Deuterostomia	CM	100 – 250	3, 423	20, 198	10.32	6153.49	28.70

[a]MT (multicelullarity type): simple or complex; UCT: number of unique cell types; GS: genome size (Megabases); GN: total number of protein-coding genes; ID (intron density): average number of introns per protein-coding gene; IS: average intron size (nucleotides); IC (intron content): the percentage of the genome size harboring intronic sequences. Refs in Lozada-Chávez et al. 2018.

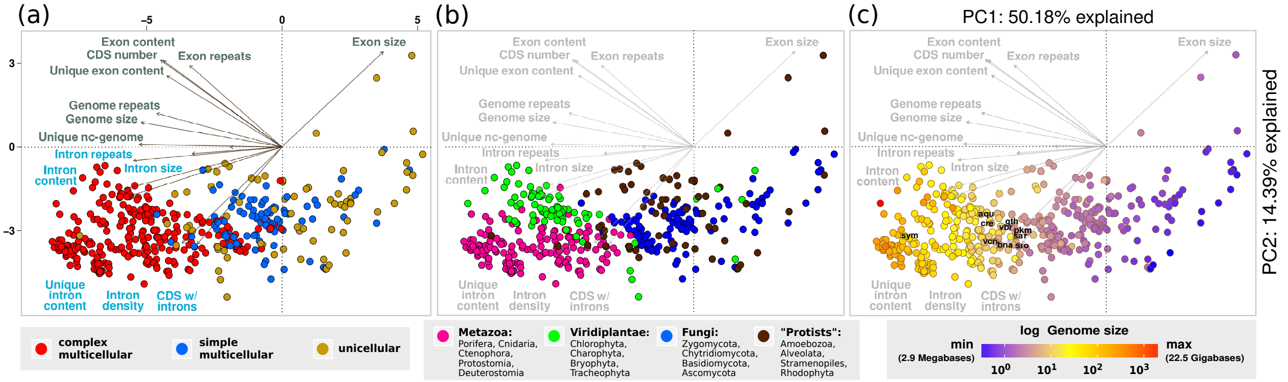

Fig. 1. Phylogenetic principal component analyses of 7 intron features, other 14 genome features, and multicellular complexity for 461 eukaryotes (Lozada-Chávez et al. 2018). Note that those genome features measuring intron–richness (in blue) cluster most complex multicellular organisms (a), independently of their taxonomy (b) and genome size (c).

Metazoa, but not in Fungi (partially due to the lack of developmental data). Furthermore, our results suggest that relatedness (in agreement with Fisher et al. 2013), intron-richness and contingent irreversibility could have influenced the likelihood of some lineages to boost the concerted development of multiple cell types, as shown by some experiments (Heyn et al. 2015). Collectively, our findings support the hypothesis that particular ncDNA traits, such as high intron-richness, were a key biological pre-condition to evolve CM on Earth. Whether or not these findings challenge the search for complex extraterrestrial life requires careful evaluation.

REFERENCES

Foucher, F., et al. 2017, Life, 7, 40

Fisher, R. M., et al. 2013, Curr Biol, 12, 1120

Grosberg, R. K. & Strathmann, R. R. 2007, Annu Rev Ecol Evol Syst, 38, 621

Heyn, P., et al. 2015, Bioessays, 37, 148

Knoll, A. H. 2011, AREPS, 39, 217

Lynch, M., et al. 2011, Annu Rev Genomics Hum Genet, 12, 347

Lozada-Chávez, I, Stadler, P. F. & Prohaska, S. J. 2011, OLEB, 41, 587

Lozada-Chávez, I, Stadler, P. F. & Prohaska, S. J. 2018, bioRxiv doi.org/10.1101/283549

Nagy, L. G. 2018, Biol Rev, 93, 1778

Niklas, K. J., et al. 2014, Acta Soc Bot Pol, 4, 337

Ratcliff, W. C., et al. 2013, NatCo, 4, 2742

Rensing, S. A. 2016, Trends Plant Sci, 7, 562