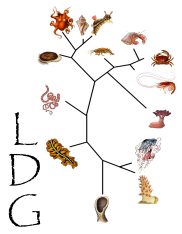
Mylena Daiana Santander

Evolução genômica dos Medusozoa (Cnidaria), com ênfase em elementos repetitivos nos Scyphozoa

Genomic evolution of Medusozoa (Cnidaria), with emphasis on repetitive elements in Scyphozoa

São Paulo
2022

instituto
de biociências



Mylena Daiana Santander

Evolução genômica dos Medusozoa (Cnidaria), com ênfase em elementos repetitivos nos Scyphozoa

Genomic evolution of Medusozoa (Cnidaria), with emphasis on repetitive elements in Scyphozoa

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo, para a obtenção de Título de Mestre em Ciências Biológicas (Biologia Genética), na Área de Biologia (Genética).

Orientadora: Sónia Cristina da Silva Andrade

Co-orientador: Maximiliano Manuel Maronna

São Paulo
2022

Ficha catalográfica elaborada pelo Serviço de Biblioteca do Instituto de Biociências da USP, com os dados fornecidos pela autora no formulário:
'https://biblioteca.ib.usp.br/ficha-catalografica/src/ficha.php'

Santander, Mylena Daiana
Evolução genômica dos Medusozoa (Cnidaria), com ênfase em elementos repetitivos nos Scyphozoa / Mylena Daiana Santander ; orientador Sónia Cristina da Silva Andrade ; coorientador Maximiliano Manuel Maronna -- São Paulo, 2022.
145 p.

Dissertação (Mestrado) -- Instituto de Biociências da Universidade de São Paulo. Programa de Pós-Graduação em Genética e Biologia Evolutiva.

1. Cnidaria. 2. Genômica. 3. Elementos repetitivos. 4. Elementos de transposição. 5. DNA satélite. I. Andrade, Sónia Cristina da Silva, orient. II. Maronna, Maximiliano Manuel, coorient. III. Título.

Bibliotecária responsável pela catalogação:
Elisabete da Cruz Neves - CRB - 8/6228

Comissão Julgadora

_____	_____
Prof(a). Dr(a).	Prof(a). Dr(a).
_____	_____
Prof(a). Dr(a).	Profa. Dra. Orientadora Sónia Cristina da Silva Andrade

A Amalia e Felipe

Acknowledgements

I am grateful to Universidade de São Paulo, Instituto de Biociências, Departamento de Genética e Biologia Evolutiva, and to the Programa de Pós-graduação em Ciências Biológicas (Biologia Genética). I am also grateful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superior for the scholarship they provided during this degree (CAPES process number 88882.377420/2019-01); to Fundação de Amparo à Pesquisa do Estado de São Paulo, whose financial support to my co-supervisor allowed me to carry out this project (FAPESP 2016/04560-9). I want to thank my supervisors Sónia Andrade and Max Maronna, for the valuable discussions and always motivating me. I want to thank André Carrara Morandini, who opened his laboratory for me. I am grateful to the past and current Darwin cluster administrators at USP, Professors Gabriel Marroig, Diogo Meyer and members of the Seção Técnica de Informática (IB), for providing and maintaining the computational resources.

To all the members of Laboratório de Diversidade Genômica and Laboratório de Estudos e Cultivos de Cnidaria, for their companionship, discussions and for welcoming me when I just arrived in São Paulo. I also want to thank Mayara Jordano and Gisele Tiseo for helping me during my first attempts at keeping species in culture. I want to especially thank some members of the sequencing project: Clarissa Molinari, Jonathan Lawley and Edgard Gamero-Mora, for maintaining the lab. cultures and participating in the DNA extraction; to Elio Castillo for teaching me the first steps in the satellite analysis; to Adrian Jaimes-Becerra, who helped me to set up software at the beginning; and Joseph Ryan for his valuable comments and teachings that contributed to this dissertation.

I want to thank my Argentinian friends, who know how to overcome the distances and despite them, have accompanied me over the past years. I am very grateful to Rafaela, who accompanied me throughout the journey and for the home we built together. I want to thank Martin for always hearing and reassuring me. Finally, I want to thank the rest of my family for always being there supporting me and motivating me to accomplish my goals.

Table of Contents

General Introduction

Historical perspective and current framework on genome size and repetitive DNA

The existence of immense variation in haploid DNA content (C-value) has been acknowledged since the middle 1960s (Cavalier-Smith, 1978; references therein). In Eukaryotes, the interspecific C-value variation is at least of five orders of magnitude, while the intraspecific variation is assumed to be almost negligible in comparison (Graur, 2016; Gregory, 2001; but see Stelzer et al., 2021). The lack of correlation between genome size and organismal or genetic complexity (i.e., number of protein-coding genes) contradicted the expectations of the scientific community at the time and remained as an apparent paradox for many years (Cavalier-Smith, 1978).

A significant proportion of eukaryotic genomes is represented by sequences that are repeated thousands or even millions of times (Charlesworth et al., 1994) whose diversity has been grouped under the concept repeatome (Y. B. Kim et al., 2014), adding to the earlier and widely extended concept, the mobilome (Siefert, 2009). The latter includes all types of sequences capable of moving within and between genomes, including transposable elements, viruses, bacteriophages and self-splicing sequences such as group II introns (Siefert, 2009). According to their pattern of genomic distribution, repetitive elements are classified into two main groups, (I) those sequences distributed in tandem and (II) those dispersed repeats (Richard et al., 2008). Tandem repeats include multigene families, minisatellites, microsatellites, and satellites (Richard et al., 2008). Dispersed repeats include transposable elements (TEs), dispersed multigene families and pseudogenes (Richard et al., 2008).

Repetitive elements are popularly associated with the inaccurate term “junk-DNA”. This concept was formally coined by Ohno (1972) to refer to the extra non-coding DNA that could explain the existence of genome size variations. Although he specifically referred to

pseudogenes (Kuska, 1998; Palazzo & Gregory, 2014), the scientific community rapidly embraced the term to indiscriminately designate different kinds of sequences that were assumed to be non-functional and neutrally evolving. This led to many contemporary debates and controversies that primarily focused on the notion of functionality, many of which are still open (e.g., ENCODE debate; Doolittle et al., 2014; Graur et al., 2013; Palazzo & Gregory, 2014).

It was soon recognized that the genome size of related taxa could vary discretly, i.e. as multiples of a basal value, or gradually (Gregory, 2001; references therein). Moreover, this variation could be caused by different interacting mechanisms, such as whole-genome duplications (WGD, i.e., polyploidy) followed by diploidization, segmental duplications, deletions, and variations in repetitive content (Blommaert, 2020; Graur, 2016; Gregory, 2001). The "C-value paradox" was outmoded by the understanding of the mechanisms behind variations in genome content, but mainly by the discovery of significant variations in the content of repetitive DNA that was considered "non-functional" (Graur, 2016; discussed below). The evolutionary processes that contribute to the apparently non-random variation in genome size are not so well known and understood and remain under the reframed concept “C-value enigma” (Gregory, 2001).

Another example of debate surrounding repetitive DNA is the concept of genetic “selfish” elements, usually applied to transposable elements but also sporadically used for satellite DNA (e.g., Charlesworth et al., 1994). Selfish elements are genomic parasites that amplify and mobilize within a host genome against individual fitness (Kidwell & Lisch, 2001). The parasite analogy, which was first coined for TEs by Orgel and Crick (1980), captures the mobilizing and amplifying mechanisms of TEs and the intragenomic selection acting on them (Kidwell & Lisch, 2001). Nevertheless, as occurred with the terms “junk”, many authors have been for and against the term “selfish” (Kidwell & Lisch, 2001; Palazzo & Gregory, 2014). The main argument against it is that it could disregard the potential benefits of repetitive elements to hosts and its many implications in genomic evolution (Kidwell & Lisch, 2001).

Some authors proposed that the relationship between TEs and their host genomes would be better described as a continuum, fluctuating from extreme parasitism to mutualism (Kidwell & Lisch, 2001). This analogy was later expanded, giving rise to the term “ecology of the genome”, where repetitive elements and their host genomes are considered species belonging to the same ecological community (Brookfield, 2005; Venner et al., 2009). Different genomic species would occupy different genomic niches, and a series of possible interactions could take part (e.g., parasitism, mutualism, neutralism) between TEs and the host sequences, and between different TE species within the same organism (Venner et al., 2009).

Finally, based on the selected-effect function at the organismal level, that is, the function a sequence was maintained for by natural selection, Graur et al. (2015) proposed four categories of genomic DNA: i. the result of selection on nucleotide sequence on “literal” ones (e.g., genes), ii. the result of selection acting on sequence distribution/structure or “indifferent” (e.g., centromeric DNA, spacers, fillers), both types considered as functional sequences; iii. “junk”, if the genomic segment is not under selection, or iv. “garbage”, if it is under non-effective purifying selection, both considered as non-functional sequences. These four categories could be transcriptionally active or not, translated or not and could change of status during evolution, e.g., junk to literal DNA (Graur et al., 2015). Thus, a careful interpretation of the diversity of repetitive elements can lead to the inclusion of different repeats in any of these four categories.

Repeatome components

Several classification schemes are available cataloging the diversity of repetitive elements and have been the subject of intense discussion, especially for TEs (Arkhipova, 2017; Charlesworth et al., 1994; Piégu et al., 2015; Storer et al., 2021; Wicker et al., 2007). This work focuses mainly on TEs and satellites because these are frequent sources of genome size variations in eukaryotes (e.g., Bosco et al., 2007; Stelzer et al., 2021; Wong et al., 2019).

Transposable elements can move through the genome (and between genomes) by different mechanisms. The first fundamental division of TEs, proposed by Finnegan (Finnegan, 1989), is based on the nature of their transposition intermediate (Wells & Feschotte, 2020) and is the base of most classification schemes. Class I retrotransposons have retro-transcribed RNA intermediates and are integrated into novel locations in a “copy-and-paste” manner. Class II transposons do not have RNA intermediates and are usually known as “DNA” transposons. Most Class II TEs transpose by a “cut-and-paste” mechanism, in which the element is cleaved and re-integrated in other chromosomal locations (Wells & Feschotte, 2020).

Transposable elements have characteristics that make them particularly difficult to classify (reviewed in Arkhipova, 2017). Briefly, such characteristics include their multiple origins and the occurrence of events that create chimeric and fragmented elements (e.g., nested insertions, recombination, arrested transposition) (Arkhipova, 2017). Existing classification schemes combine different cladistic, mechanistic, and structural criteria to deal with TE diversity and therefore differ in assigning nested hierarchies such as class, order, superfamily, family or subfamily of TEs (Storer et al., 2021). For practical and operational reasons, a family of TEs is defined by a 80-80-80 rule (Wicker et al., 2007): a family includes sequences (>80pb) that show 80% similarity in 80% of their sequence.

Satellites (satDNA) are highly repetitive monomeric sequences organized in tandem, forming blocks in centromeres and heterochromatic regions (reviewed in Plohl et al., 2012). They are mainly distinguished from other tandem repeats such as micro and minisatellites by the size of their arrays, which is two orders of magnitude higher than minisatellites (Richard et al., 2008). Despite its structural roles in chromosome structure, satellite DNA shows rapid evolutionary changes due to different processes involved in its origin, dispersion and homogenization, leading to intra and interspecific variations (Garrido-Ramos, 2017; Lower et al., 2018; Plohl et al., 2012). A formal classification scheme for satDNA is still lacking, but its diversity is usually classified into families and superfamilies depending on sequence similarity (e.g., Ruiz-Ruano et al., 2016).

SatDNA and TEs are traditionally considered differentiated elements (e.g., Charlesworth et al., 1994), but there is a growing amount of evidence that these elements are closely interrelated and constitute a complex network that affects genomic architecture (reviewed in Meštrović et al., 2015; Paço et al., 2019). Both utilize similar chromosomal niches, satDNA can be originated by the activity of TEs and also mobilized by them, and both can have copies distributed in dispersed and clustered manners (Paço et al., 2019).

The different molecular mechanisms related to TEs and satDNAs biology turn them into potential mutagenic agents, producing genetic restructuring at genomic or chromosomal scales and alterations in gene structure and expression (Biscotti et al., 2015; Feschotte, 2008; Warren et al., 2015). There is growing evidence that TEs can be recruited into novel functions or genes (Cosby et al., 2021; Feschotte, 2008). SatDNA has been shown to play an active role in the formation and maintenance of chromatin structure, affecting genomic integrity and stability (Biscotti et al., 2015; Garrido-Ramos, 2017). Finally, the repeatome is currently considered a potential source of evolutionary novelties that impact on the macroevolution of organisms (Y. B. Kim et al., 2014; Warren et al., 2015). A broader exploration of the great diversity of repetitive elements along the tree of life would have a central role in developing of our understanding of the evolution of eukaryotic genomes and their consequences.

TEs: evolutionary mechanisms and related processes

The mobile capacity of TEs would favor their increase in copy number unless constrained by natural selection. In summary, three forces affect the success of TE spread (Wells & Feschotte, 2020): i. the rate of transposition, ii. the rate of fixation of insertions, and iii. the rate of deletion and erosion. At the same time, these three forces are influenced by populational, genetic, and environmental factors that interplay during the lifecycle of a TE (Wells & Feschotte, 2020). Different mechanisms of TE control and silencing have evolved that affect the rate of transposition and deletion. In animals, one of the main mechanisms of

TE silencing is the PIWI and pi-RNA pathway that prevents TE mobilization by epigenetic marks and direct degradation of TE-derived RNAs (Yamanaka et al., 2014).

On the other hand of this intragenomic “race”, TEs tend to overcome host silencing mechanisms by random mutation, insertion preferences and lateral transmission (Wells & Feschotte, 2020). Mobile elements are often considered sexually-transmitted, as sexual reproduction and recombination can allow TEs to escape silencing by invading novel genomic backgrounds. In this context, asexuality can lead to the selection of reduced insertion rates and effective silencing mechanisms, therefore reducing TE-load over time (Boutin et al., 2012; Glémin et al., 2019). An opposite hypothesis proposes that TE-load would increase in asexuals due to a decrease in the effectiveness of purifying selection on linked traits but would eventually be lost by a Muller’s “ratchet” like mechanism (Glémin et al., 2019 and references therein).

To understand the accumulation/loss of repetitive elements, we also have to recognize processes acting on a higher level, that is, external forces affecting genome size change. Both selective and neutral hypotheses have been proposed to address this issue and have gained support in different lines of evidence and different groups of organisms (reviewed in Blommaert, 2020).

Genomic size can be positively correlated to cell and body size and negatively to developmental and metabolic rates, among other traits (Alfsnes et al., 2017; Beaudreau et al., 2021; but see Gardner et al., 2020). Two main selective hypotheses were proposed to explain these correlations. The ***Nucleotypic Hypothesis*** establishes that genome size is under selection, as DNA quantity is expected to affect phenotype directly (Bennett, 1972). According to the ***Nucleoskeletal hypothesis***, genome size is a sub-product of selection acting on other features which depend on DNA content acting as a structural component of the nucleus (Cavalier-Smith, 1978).

On the other hand, two (nearly)neutral hypotheses were recently proposed to explain genome size variation. Following the ***Mutational Equilibrium Hypothesis***, genome size would be determined by the balance between the rate of “junk” DNA accumulation and its

gradual loss by drift (Petrov, 2002). In other words, genome size would result from the equilibrium between insertion and deletion rates. On the contrary, the **Mutational Hazard Hypothesis** considers population size as the main factor driving genome size evolution (Lynch & Conery, 2003; but see Roddy et al., 2021). In populations with small effective population sizes, slightly deleterious insertions could be fixed by drift, and these populations would tend to accumulate DNA over time (Lynch & Conery, 2003). The contrary would be true in sufficiently large populations, in which slightly deleterious insertions or duplications would be filtered out by natural selection (Lynch & Conery, 2003).

Ultimately, a single factor would not explain the eukaryotic genome size range and the great diversity of repetitive elements. Several historical, ecological, and genetic factors (or a combination of them) can have different importance in independent lineages (Alfsnes et al., 2017; Wells & Feschotte, 2020).

Medusozoans and their genomic organization

The cnidarian clade Medusozoa (4,073 current species) is distinguished by a series of characteristics, including the presence of metagenic life cycles with three differentiated stages (i.e., polyp, medusae and larvae), linear mitochondria, varied reproduction strategies and demographic phenomena known as blooms and swarms (Arai, 1997; Dawson & Hamner, 2009; Kayal et al., 2018; Morandini et al., 2016). Medusozoa is composed of four different classes, of which Hydrozoa and Scyphozoa are the most species-rich, with ca. 3,752 and 224 contemporary species respectively, followed by the less specious classes Cubozoa (48 spp.) and Staurozoa (49 spp.; *World Register of Marine Species*, 2022).

According to a recent and robust phylogenetic hypothesis of Cnidaria (Kayal et al., 2018), two main Medusozoa lineages were reconstructed: Hydrozoa and Acraspeda, the latter composed of the remaining classes as follows: (Staurozoa(Scyphozoa, Cubozoa)) (Figure 1A). The clade composed of Scyphozoa and Cubozoa, known as Rhopaliophora, stands out for the predominance of the life cycle stage known as jellyfish or medusae (pelagic sexually active stage), of non-colonial polyps and the presence of sensitive

structures called rhopalia (Kayal et al., 2018). Scyphozoa is further divided into two main clades: the subclasses Coronamedusae (7 families, 13 genera and 59 spp.) and Discomedusae (15 families, 48 genera and 165 spp.; Bayha et al., 2010; Kayal et al., 2018; *World Register of Marine Species*, 2022).

Despite the relatively small number of Scyphozoa species and their apparent simplicity, it harbors variable ecological, biological, and populational features that can potentially impact genome dynamics and evolution (discussed below). This, together with the fact that many species can be maintained in laboratory conditions (Jarms et al., 2002), make them a proper biological system for testing different evolutionary hypotheses such as the evolution of complex life cycles and the establishment of symbiosis, among others.

The metagenic life cycles imply the existence of four stages that succeed during a scyphozoan lifetime and differ in body plan and environment (Djeghri et al., 2019; Figure 1B): the planula larvae (short-lived, free-swimming), the polyp (benthic, sessile, asexually-reproducing), the ephyrae (immature jellyfish) and the medusae (pelagic free-swimming sexually-reproducing). The evolution of species with complex life cycles would be affected by developmental and genetic constraints because the same genome would be under different selective regimes at the stages presenting different lifestyles (e.g., solitary/colonial, planktic/benthic; Albecker et al., 2021; Moran, 1994).

These contrasting patterns between stages can also be found between species that have lost or reduced a life stage, for example becoming holoplantic or holobenthic. Such reductions include the Coronatae species *Periphylla peryphilla* (direct development without a polyp stage; Jarms et al., 1999), and *Thecoscyphus zibrowii* (jellyfish reduced to a reproductive structure called egg sack, which sometimes remains fixed to the polyp; Sötje & Jarms, 2009). The Medusozoa general life cycle pattern shows remarkable plasticity, as many other deviations from it were described, such as life-cycle reversions (Morandini et al., 2016).

Several mechanisms of asexual reproduction exist in Medusozoa besides metamorphosis and metagenesis: podocyst formation, lateral budding, stolons, and

regeneration, (Schiariti et al., 2014). Species can asexualy reproduce by one (mono-mode) or several (multi-mode) of the aforementioned strategies. For example, both multi-mode and mono-mode species have been identified in Scyphozoa, of which multi-mode presented the highest rates of polyp production (Schiariti et al., 2014). Three main topics are important regarding reproductive modes: (I) it is suggested that rapid propagation by asexual strategies can impact population size and predispose species to bloom (Schiariti et al., 2014); (II) the effect of different reproductive strategies in the evolution of genome size; and (III) asexual propagation can pass on somatic mutations to new individuals. If we also consider the potential perennial nature of polyps (Morandini et al., 2016) and the hypothesis of a lack of germline in Cnidaria (Watanabe et al., 2009), scyphozoans are suitable for studying the contribution of somatic mutations in animal evolution.

In terms of ecological interactions, the establishment of the photosymbiosis with eukaryotic zooxanthellae would have happened at least seven times independently in Medusozoa (Djeghri et al., 2019). In Scyphozoa, species from two Coronamedusae genera and 12 Discomedusae genera harbor zooxanthellae, some of them facultatively (reviewed in Djeghri et al., 2019). Zooxanthellae are vertically transmitted during asexual reproduction and strobilation or taken up by polyps after larvae metamorphosis (Djeghri et al., 2019). The nutritional dependence on symbiosis is variable among species, and it is suggested that photosynthetic contribution to nutrition would be higher in the medusae than in the polyp stage (Djeghri et al., 2019). Establishing the symbiotic relationship is known to have profound impacts on the genomes of both the symbiont and the host, including extensive reduction of the first and acquisition of laterally transferred genes by the second, among other effects (Sloan et al., 2014; and references therein).

In recent years, significant advances have taken place regarding Medusozoa genomics. With multiple genomes being sequenced and assembled, different topics regarding Medusozoa-centric questions are being addressed, such as the emergence of the medusae and complex venom sets. Nevertheless, compared with other groups of animals, our knowledge of their genomic organization is still scant. This is especially evident when

considering repetitive elements that, with very few exceptions, were not the main objective of most genome projects and reported information was very limited.

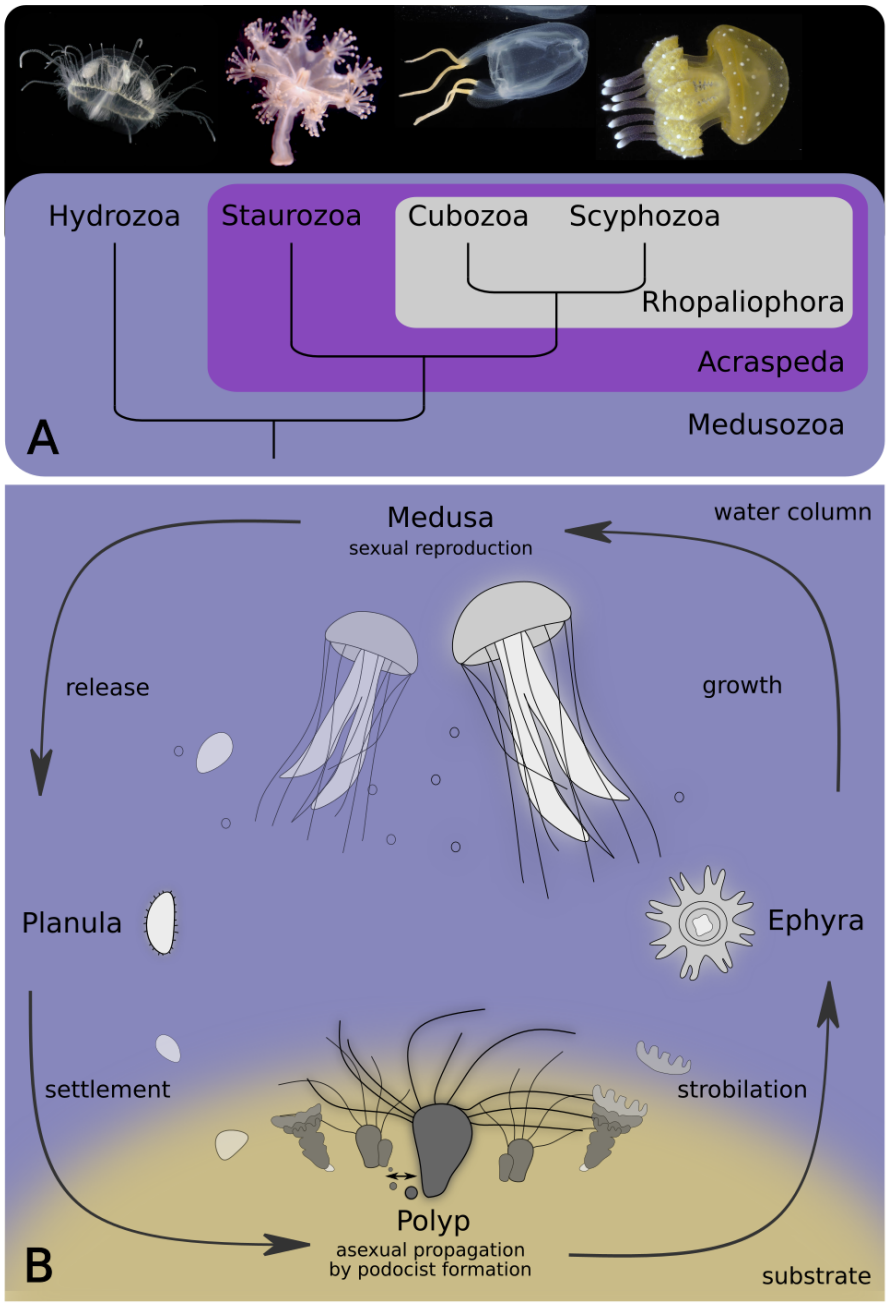


Figure 1 - Medusozoa phylogeny, main clade representatives, and lifecycle. A. Phylogeny and clade representatives. The tree topology was retrieved from previous phylogenetic analyses (Kayal et al., 2018; and references therein). Species names are as follows (from left to right): *Craspedacusta sowerbii, Haliclystus sanjuanensis, Tamoya haplonema* and *Phyllorhiza punctata.* Except for the *C. sowerbii* photograph that Marta Chiodin kindly provided; all other photographs were obtained from Cifonauta (Migotto & Vellutini, n.d.) and were captured by Alvaro E. Migotto (Hydrozoa and Cubozoa) and André Carrara Morandini (Scyphozoa). B. Schematic representation of Medusozoa complex life cycle, exemplified by the scyphozoan *Chrysaora lactea* (Morandini et al., 2004). General life cycle variations were described for almost all of the stages in the different clades. For example, the ephyrae are supposed to be exclusive of Scyphozoa, and strobilation is replaced by metamorphosis in Cubozoa and Staurozoa (Straehler-Pohl & Jarms, 2022).

Hydra is the most extensively studied from this point of view, where studies have focused on the determination of the repetitive elements responsible for the genome size expansion (Wong et al., 2019), on microsatellite marker description and dynamics (Ruiz-Ramos & Baums, 2014; Schenkelaars et al., 2020), and mechanisms of TE repression (Teefy et al., 2020; Ying et al., 2022). Two main pathways of TE repression were identified: the PIWI pi-RNA pathway and DNA methylation. In the first case, PIWI–piRNA complexes that produce the cytoplasmic degradation TE-derived RNAs repressed transposition in *Hydra vulgaris* (Teefy et al., 2020). Unlike other metazoans, where the TE repression activity of this pathway is restricted to the germ line, PIWI-piRNAS actively degrade TEs in somatic cells in *Hydra* (Teefy et al., 2020). In the second case, CpG methylation targeted young TE-insertions and declined in more divergent ones (i.e., older copies), suggesting a role in TE immobilization (Ying et al., 2022). In the remaining Medusozoa, transposable elements and microsatellites have been characterized to some extent (i.e., TE annotation in most published genomes), but satellite DNA was estimated with non-specialized tools in two scyphozoan species with no further characterization (Gold et al., 2019; H.-M. Kim et al., 2019).

This dissertation aims to contribute to the knowledge of genome content and organization in Medusozoa, with emphasis on Scyphozoa. For that, this work was organized in two independent but interconnected chapters related to this thematic. The first chapter, entitled “The state of Medusozoa genomics: current evidence and future challenges”, is focused on reviewing all available information about Medusozoa genomes, that is, from karyotypes and genome size estimations, to the recent advances in sequencing projects. We also addressed potential standardization problems that could potentially hinder novel discoveries.

In the second chapter, entitled “Jellyfish repeatomes uncover hidden satellite diversity and provide insights into the C-value enigma”, we presented novel genomic information for several scyphozoan species representing its two main clades and studied

their repeatomes in a comparative framework for all available datasets to understand the evolution of repetitive elements in the group and its putative association to genome size and the C-value enigma. Finally, general conclusions were presented in the last part of the dissertation.

References

Albecker, M. A., Wilkins, L. G. E., Krueger-Hadfield, S. A., Bashevkin, S. M., Hahn, M. W., Hare, M. P., Kindsvater, H. K., Sewell, M. A., Lotterhos, K. E., & Reitzel, A. M. (2021). Does a complex life cycle affect adaptation to environmental change? Genome-informed insights for characterizing selection across complex life cycle. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1964), 20212122. https://doi.org/10.1098/rspb.2021.2122

Alfsnes, K., Leinaas, H. P., & Hessen, D. O. (2017). Genome size in arthropods; different roles of phylogeny, habitat and life history in insects and crustaceans. *Ecology and Evolution*, *7*(15), 5939–5947. https://doi.org/10.1002/ece3.3163

Arai, M. N. (1997). A functional biology of Scyphozoa. Chapman & Hall.

Arkhipova, I. R. (2017). Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. *Mobile DNA*, *8*(1), 19. https://doi.org/10.1186/s13100-017-0103-2

Bayha, K. M., Dawson, M. N., Collins, A. G., Barbeitos, M. S., & Haddock, S. H. D. (2010). Evolutionary relationships among scyphozoan jellyfish families based on complete taxon sampling and phylogenetic analyses of 18S and 28S ribosomal DNA. *Integrative and Comparative Biology*, *50*(3), 436–455. https://doi.org/10.1093/icb/icq074

Beaudreau, N., Massamba-N'Siala, G., Belzile, C., Calosi, P., & Dufresne, F. (2021). Life-history traits display strong associations to genome size in annelids. *Hydrobiologia*, *848*(4), 799–810. https://doi.org/10.1007/s10750-020-04477-7

Bennett, M. D. (1972). Nuclear DNA content and minimum generation time in herbaceous plants. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *181*(1063), 109–135. https://doi.org/10.1098/rspb.1972.0042

Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Research*, *23*(3), 415–420. https://doi.org/10.1007/s10577-015-9499-z

Blommaert, J. (2020). Genome size evolution: Towards new model systems for old questions. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1933), 20201441. https://doi.org/10.1098/rspb.2020.1441

Bosco, G., Campbell, P., Leiva-Neto, J. T., & Markow, T. A. (2007). Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics*, *177*(3), 1277–1290. https://doi.org/10.1534/genetics.107.075069

Boutin, T. S., Le Rouzic, A., & Capy, P. (2012). How does selfing affect the dynamics of selfish transposable elements? *Mobile DNA*, *3*(1), 5. https://doi.org/10.1186/1759-8753-3-5

Brookfield, J. F. Y. (2005). The ecology of the genome—Mobile DNA elements and their

hosts. *Nature Reviews Genetics*, *6*(2), 128–136. https://doi.org/10.1038/nrg1524

Cavalier-Smith, T. (1978). Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *Journal of Cell Science*, *34*(1), 247–278. https://doi.org/10.1242/jcs.34.1.247

Charlesworth, B., Sniegowski, P., & Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, *371*(6494), 215–220. https://doi.org/10.1038/371215a0

Cosby, R. L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E. J., & Feschotte, C. (2021). Recurrent evolution of vertebrate transcription factors by transposase capture. *Science*, *371*(6531). https://doi.org/10.1126/science.abc6405

Dawson, M. N., & Hamner, W. M. (2009). A character-based analysis of the evolution of jellyfish blooms: adaptation and exaptation. *Hydrobiologia*, *616*(1), 193–215. https://doi.org/10.1007/s10750-008-9591-x

Djeghri, N., Pondaven, P., Stibor, H., & Dawson, M. N. (2019). Review of the diversity, traits, and ecology of zooxanthellate jellyfishes. *Marine Biology*, *166*(11), 147. https://doi.org/10.1007/s00227-019-3581-6

Doolittle, W. F., Brunet, T. D. P., Linquist, S., & Gregory, T. R. (2014). Distinguishing between “Function” and “Effect” in genome biology. *Genome Biology and Evolution*, *6*(5), 1234–1237. https://doi.org/10.1093/gbe/evu098

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*, *9*(5), 397–405. https://doi.org/10.1038/nrg2337

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends in Genetics*, *5*(4), 103–107. https://doi.org/10.1016/0168-9525(89)90039-5

Gardner, J. D., Laurin, M., & Organ, C. L. (2020). The relationship between genome size and metabolic rate in extant vertebrates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1793), 20190146. https://doi.org/10.1098/rstb.2019.0146

Garrido-Ramos, M. A. (2017). Satellite DNA: An evolving topic. *Genes*, *8*(9), 230. https://doi.org/10.3390/genes8090230

Glémin, S., François, C. M., & Galtier, N. (2019). Genome evolution in outcrossing vs. selfing vs. asexual species. In M. Anisimova (Ed.), Evolutionary Genomics. Statistical and Computational Methods. Humana Press. https://doi.org/10.1007/978-1-4939-9074-0_11

Gold, D. A., Katsuki, T., Li, Y., Yan, X., Regulski, M., Ibberson, D., Holstein, T., Steele, R. E., Jacobs, D. K., & Greenspan, R. J. (2019). The genome of the jellyfish *Aurelia* and the evolution of animal complexity. *Nature Ecology & Evolution*, *3*(1), 96–104. https://doi.org/10.1038/s41559-018-0719-8

Graur, D. (2016). Molecular and genome evolution. Sinauer associates.

Graur, D., Zheng, Y., & Azevedo, R. B. R. (2015). An evolutionary classification of genomic function. *Genome Biology and Evolution*, *7*(3), 642–645. https://doi.org/10.1093/gbe/evv021

Graur, D., Zheng, Y., Price, N., Azevedo, R. B. R., Zufall, R. A., & Elhaik, E. (2013). On the immortality of television sets: “Function” in the human genome according to the evolution-free gospel of ENCODE. *Genome Biology and Evolution*, *5*(3), 578–590. https://doi.org/10.1093/gbe/evt028

Gregory, T. R. (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews*, *76*(1), 65–101. https://doi.org/10.1111/j.1469-185X.2000.tb00059.x

Jarms, G., Båmstedt, U., Tiemann, H., Martinussen, M. B., Fosså, J. H., & Høisœter, T.

(1999). The holopelagic life cycle of the deep-sea medusa *Periphylla periphylla* (Scyphozoa, Coronatae). *Sarsia*, *84*(1), 55–65.

Jarms, G., Morandini, A., & Da Silveira, F. (2002). Cultivation of polyps and medusae of Coronatae (Cnidaria, Scyphozoa) with a brief review of important characters. *Helgoland Marine Research*, *56*(3), 203–210. https://doi.org/10.1007/s10152-002-0113-3

Kayal, E., Bentlage, B., Sabrina Pankey, M., Ohdera, A. H., Medina, M., Plachetzki, D. C., Collins, A. G., & Ryan, J. F. (2018). Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. *BMC Evolutionary Biology*, *18*(1), 68. https://doi.org/10.1186/s12862-018-1142-0

Kidwell, M. G., & Lisch, D. R. (2001). Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution*, *55*(1), 1–24. https://doi.org/10.1111/j.0014-3820.2001.tb01268.x

Kim, H.-M., Weber, J. A., Lee, N., Park, S. G., Cho, Y. S., Bhak, Y., Lee, N., Jeon, Y., Jeon, S., Luria, V., Karger, A., Kirschner, M. W., Jo, Y. J., Woo, S., Shin, K., Chung, O., Ryu, J.-C., Yim, H.-S., Lee, J.-H., … Yum, S. (2019). The genome of the giant Nomura's jellyfish sheds light on the early evolution of active predation. *BMC Biology*, *17*(1), 28. https://doi.org/10.1186/s12915-019-0643-7

Kim, Y. B., Oh, J. H., McIver, L. J., Rashkovetsky, E., Michalak, K., Garner, H. R., Kang, L., Nevo, E., Korol, A. B., & Michalak, P. (2014). Divergence of *Drosophila melanogaster* repeatomes in response to a sharp microclimate contrast in Evolution Canyon, Israel. *Proceedings of the National Academy of Sciences*, *111*(29), 10630–10635. https://doi.org/10.1073/pnas.1410372111

Kuska, B. (1998). Should scientists scrap the notion of junk DNA? *Journal of the National Cancer Institute*, *90*(14), 1032–1033. https://doi.org/10.1093/jnci/90.14.1032

Lower, S. S., McGurk, M. P., Clark, A. G., & Barbash, D. A. (2018). Satellite DNA evolution: old ideas, new approaches. *Current Opinion in Genetics & Development*, *49*, 70–78. https://doi.org/10.1016/j.gde.2018.03.003

Lynch, M., & Conery, J. S. (2003). The origins of genome complexity. *Science*, *302*(5649), 1401–1404. https://doi.org/10.1126/science.1089370

Meštrović, N., Mravinac, B., Pavlek, M., Vojvoda Zeljko, T., Satovic, E., & Plohl, M. (2015). Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosome Research*, *23*, 583–596. https://doi.org/10.1007/s10577-015-9483-7

Migotto, A. E., & Vellutini, B. C. (n.d.). Cifonauta: banco de imagens de biologia marinha. Retrieved February 2, 2022, from http://cifonauta.cebimar.usp.br/

Moran, N. A. (1994). Adaptation and constraint in the complex life cycles of animals. *Annual Review of Ecology and Systematics*, *25*(1), 573–600. https://doi.org/10.1146/annurev.es.25.110194.003041

Morandini, A. C., Schiariti, A., Stampar, S. N., Maronna, M. M., Straehler-Pohl, I., & Marques, A. C. (2016). Succession of generations is still the general paradigm for scyphozoan life cycles. *Bulletin of Marine Science*, *92*(3), 343–351. https://doi.org/10.5343/bms.2016.1018

Morandini, A. C., Silveira, F. L. da, & Jarms, G. (2004). The life cycle of *Chrysaora lactea* Eschscholtz, 1829 (Cnidaria, Scyphozoa) with notes on the scyphistoma stage of three other species. *Hydrobiologia*, *530*(1), 347–354. https://doi.org/10.1007/s10750-004-2694-0

Ohno, S. (1972). So much "junk" DNA in our genome. Evolution of Genetic Systems, Brookhaven Symposium in Biology, *36*, 366–370.

Orgel, L. E., & Crick, F. H. C. (1980). Selfish DNA: the ultimate parasite. *Nature*, *284*(5757), 604–607. https://doi.org/10.1038/284604a0

Paço, A., Freitas, R., & Vieira-da-Silva, A. (2019). Conversion of DNA Sequences: from a transposable element to a tandem repeat or to a gene. *Genes*, *10*(12), 1014. https://doi.org/10.3390/genes10121014

Palazzo, A. F., & Gregory, T. R. (2014). The case for Junk DNA. *PLOS Genetics*, *10*(5), e1004351. https://doi.org/10.1371/journal.pgen.1004351

Petrov, D. A. (2002). Mutational Equilibrium Model of genome size evolution. *Theoretical Population Biology*, *61*(4), 531–544. https://doi.org/10.1006/tpbi.2002.1605

Piégu, B., Bire, S., Arensburger, P., & Bigot, Y. (2015). A survey of transposable element classification systems – A call for a fundamental update to meet the challenge of their diversity and complexity. *Molecular Phylogenetics and Evolution*, *86*, 90–109. https://doi.org/10.1016/j.ympev.2015.03.009

Plohl, M., Meštrović, N., & Mravinac, B. (2012). Satellite DNA evolution. *Repetitive DNA*, 7, 126–152. https://doi.org/10.1159/000337122

Richard, G.-F., Kerrest, A., & Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiology and Molecular Biology Reviews*, *72*(4), 686–727. https://doi.org/10.1128/MMBR.00011-08

Roddy, A. B., Alvarez-Ponce, D., & Roy, S. W. (2021). Mammals with small populations do not exhibit larger genomes. *Molecular Biology and Evolution*, *38*(9), 3737–3741. https://doi.org/10.1093/molbev/msab142

Ruiz-Ramos, D. V., & Baums, I. B. (2014). Microsatellite abundance across the Anthozoa and Hydrozoa in the phylum Cnidaria. *BMC Genomics*, *15*, 939. https://doi.org/10.1186/1471-2164-15-939

Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., & Camacho, J. P. M. (2016). High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Scientific Reports*, *6*(1), 28333. https://doi.org/10.1038/srep28333

Schenkelaars, Q., Perez-Cortes, D., Perruchoud, C., & Galliot, B. (2020). The polymorphism of *Hydra* microsatellite sequences provides strain-specific signatures. *PLOS ONE*, *15*(9), e0230547. https://doi.org/10.1371/journal.pone.0230547

Schiariti, A., Morandini, A., Jarms, G., von Glehn Paes, R., Franke, S., & Mianzan, H. (2014). Asexual reproduction strategies and blooming potential in Scyphozoa. *Marine Ecology Progress Series*, *510*, 241–253. https://doi.org/10.3354/meps10798

Siefert, J. L. (2009). Defining the Mobilome. In M. B. Gogarten, J. P. Gogarten, & L. C. Olendzenski (Eds.), Horizontal Gene Transfer. Humana Press. https://doi.org/10.1007/978-1-60327-853-9_2

Sloan, D. B., Nakabachi, A., Richards, S., Qu, J., Murali, S. C., Gibbs, R. A., & Moran, N. A. (2014). Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Molecular Biology and Evolution*, *31*(4), 857–871. https://doi.org/10.1093/molbev/msu004

Sötje, I., & Jarms, G. (2009). Derivation of the reduced life cycle of *Thecoscyphus zibrowii* Werner, 1984 (Cnidaria, Scyphozoa). *Marine Biology*, *156*(11), 2331–2341.

Stelzer, C. P., Blommaert, J., Waldvogel, A. M., Pichler, M., Hecox-Lea, B., & Mark Welch, D. B. (2021). Comparative analysis reveals within-population genome size variation in a rotifer is driven by large genomic elements with highly abundant satellite DNA repeat elements. *BMC Biology*, *19*(1), 206. https://doi.org/10.1186/s12915-021-01134-w

Storer, J., Hubley, R., Rosen, J., Wheeler, T. J., & Smit, A. F. (2021). The Dfam community resource of transposable element families, sequence models, and genome

annotations. *Mobile DNA*, *12*(1), 2. https://doi.org/10.1186/s13100-020-00230-y

Straehler-Pohl, I., & Jarms, G. (2022). Back to the roots, Part 1—Early life cycle data of Rhopaliophora (Scyphozoa, Cubozoa and Staurozoa). *Plankton and Benthos Research*, *17*(1), 1–33. https://doi.org/10.3800/pbr.17.1

Teefy, B. B., Siebert, S., Cazet, J. F., Lin, H., & Juliano, C. E. (2020). PIWI–piRNA pathway-mediated transposable element repression in *Hydra* somatic stem cells. *RNA*, *26*(5), 550-563. https://doi.org/10.1261/rna.072835.119

Venner, S., Feschotte, C., & Biémont, C. (2009). Dynamics of transposable elements: towards a community ecology of the genome. *Trends in Genetics*, *25*(7), 317–323. https://doi.org/10.1016/j.tig.2009.05.003

Warren, I. A., Naville, M., Chalopin, D., Levin, P., Berger, C. S., Galiana, D., & Volff, J.-N. (2015). Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. *Chromosome Research*, *23*(3), 505–531. https://doi.org/10.1007/s10577-015-9493-5

Watanabe, H., Hoang, V. T., Mättner, R., & Holstein, T. W. (2009). Immortality and the base of multicellular life: lessons from cnidarian stem cells. *Seminars in Cell & Developmental Biology*, *20*(9), 1114–1125. https://doi.org/10.1016/j.semcdb.2009.09.008

Wells, J. N., & Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annual Review of Genetics*, *54*(1), 539–561. https://doi.org/10.1146/annurev-genet-040620-022145

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(12), 973–982. https://doi.org/10.1038/nrg2165

Wong, W. Y., Simakov, O., Bridge, D. M., Cartwright, P., Bellantuono, A. J., Kuhn, A., Holstein, T. W., David, C. N., Steele, R. E., & Martínez, D. E. (2019). Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus *Hydra*. *Proceedings of the National Academy of Sciences*, *116*(46), 22915–22917. https://doi.org/10.1073/pnas.1910106116

WoRMS - World Register of Marine Species. (2022). Retrieved November 4, 2021, from https://www.marinespecies.org. https://www.doi.org/10.14284/170

Yamanaka, S., Siomi, M. C., & Siomi, H. (2014). piRNA clusters and open chromatin structure. *Mobile DNA*, *5*(1), 22. https://doi.org/10.1186/1759-8753-5-22

Ying, H., Hayward, D. C., Klimovich, A., Bosch, T. C. G., Baldassarre, L., Neeman, T., Forêt, S., Huttley, G., Reitzel, A. M., Fraune, S., Ball, E. E., & Miller, D. J. (2022). The role of DNA methylation in genome defense in Cnidaria and other invertebrates. *Molecular Biology and Evolution*, *39*(2), msac018. https://doi.org/10.1093/molbev/msac018

Abstract

A significant portion of the eukaryotic genome comprises a great diversity of repetitive elements, collectively known as the repeatome. Despite being previously considered expendable, it has gained revived importance at the present due to its potential as a source of evolutionary novelties. Medusozoa is a widely distributed ancient lineage that harbors one-third of Cnidaria diversity. This clade is characterized by the existence of complex life cycles that, together with body plans, show marked plasticity. Like other early-diverging Metazoa, the information related to their genomic organization and their repeatome is scarce, a gap that we aimed to fill. First, we reviewed different sources of genetic and genomic information, including cytogenetic records and high-throughput sequencing (HTS) projects. We highlighted a lack of standardization in genomic projects which can potentially affect evolutionary inferences. Then, we restricted our analysis to the class Scyphozoa which harbors ca. 242 species. The availability of genomic data coupled with variable biological characteristics and the relatively small and repeat-rich genomes of scyphozoans make them suitable study systems for the study of the evolution of the repeatome. Based on the use of HTS datasets, we carried out a comparative analysis of the repetitive content of 12 species of Scyphozoa that showed a 11-fold genome size variation. For the first time, we carried out a detailed characterization of the satellite content in Scyphozoa and found high variation in the number of satellite families per species and their features (abundance, GC%, monomer length). Moreover, we observed that larger genomes had higher percentages of transposable elements (TEs: ~20-35%; satellites: ~3-10%), while smaller genomes were generally dominated by satellites or showed similar proportions of both elements (TEs: ~2-25%; satellites: ~1-22%). We propose general trends regarding Scyphozoa repeatome and the implications of the abundance and diversity of different transposable elements in the evolution of their genome sizes.

Resumo

Uma parcela significativa dos genomas de eucariotos é composta por uma grande diversidade de elementos repetitivos, conhecidos coletivamente como repitoma. Apesar de anteriormente considerado dispensável, o repitoma ganhou importância revivida na atualidade devido ao seu potencial como fonte de novidades evolutivas. Medusozoa é uma linhagem antiga, amplamente distribuída que abriga um terço da diversidade de Cnidaria. Este clado é caracterizado pela existência de ciclos de vida complexos que, juntamente com os planos corporais, apresentam marcada plasticidade. Como em outros grupos de metazoários basais, as informações relacionadas à sua organização genômica e seu repitoma são escassas, uma lacuna que procuramos preencher. Primeiro, revisamos diferentes fontes de informações genéticas e genômicas, incluindo registros citogenéticos e projetos de sequenciamento de alto rendimento (HTS). Destacamos a falta de padronização em projetos genômicos que podem afetar potencialmente as inferências evolutivas. Em seguida, restringimos nossa análise à classe Scyphozoa que abriga ca. 242 espécies. A disponibilidade de dados genômicos de cifozoários combinada com as suas características biológicas variáveis e genomas relativamente pequenos e ricos em DNA repetitivo, os tornam sistemas de estudo adequados para o estudo da evolução do repitoma. Com base no uso de conjuntos de dados HTS, realizamos uma análise comparativa do conteúdo repetitivo de 12 espécies de Scyphozoa que apresentaram uma variação de 11 vezes no tamanho do genoma. Pela primeira vez, realizamos uma caracterização detalhada do conteúdo de satélites em Scyphozoa e encontramos alta variação no número de famílias de satélites por espécie e suas características (abundância, GC%, comprimento do monômero). Além disso, observamos que genomas maiores tinham porcentagens mais altas de elementos de transposição (TEs: ~20-35%; satellites: ~3-10%), enquanto genomas menores eram geralmente dominados por satélites ou mostravam proporções semelhantes de ambos os elementos (TEs: ~2 -25%; satellites: ~1-22%). Propomos tendências gerais

sobre o repitoma de Scyphozoa e as implicações da abundância e diversidade de diferentes elementos de transposição na evolução de seus genomas.